COK-H01163-95-P004926

THE AMERICAN ECONOMIC REVIEW

95

391
r 001

P 4926

Editor

ORLEY ASHENFELTER

Co-Editors

ROBERT H. HAVEMAN
JOHN B. TAYLOR
HAL R. VARIAN

Production Editor

CLAIRE H. COMISKEY

Board of Editors

GEORGE A. AKERLOF
JACOB A. FRENKEL
CLAUDIA D. GOLDIN
JO ANNA GRAY
GEORGE E. JOHNS
JOHN F. KENN
MERVYN A. KING
BENNETT T. McCALL
MAURICE OBSTFELD
EDGAR O. OLSEN
ROBERT H. PORTER
RICHARD ROLL
ALVIN E. ROTH
STEVEN SHAVELL
JOHN B. SHOVEN
KENNETH J. SINGLETON

71
001

March 1988

VOLUME 78, NUMBER 1

Articles

Family Economics and Macro Behavior *Gary S. Becker* 1

The Effects of Fiscal Policies When Incomes Are Uncertain: A Contradiction to Ricardian Equivalence *Martin Feldstein* 14

Asset Prices, Commodity Prices, and Money: A General Equilibrium, Rational Expectations Model *Glenn W. Boyle and Leslie Young* 24

The Effects of Commodity Price Stabilization Programs *Mario J. Miranda and Peter G. Helmberger* 46

Counterfeit-Product Trade *Gene M. Grossman and Carl Shapiro* 59

Foreign-Owned Land *Jonathan Eaton* 76

New Estimates of Quality of Life in Urban Areas *Glenn C. Blomquist, Mark C. Berger, and John P. Hoehn* 89

The Division of Markets Is Limited by the Extent of Liquidity *Nicholas Economides and Aloysius Siow* 108

On the Strategic Choice of Spatial Price Policy *Jacques-Francois Thisse and Xavier Vives* 122

Will Share Contracts Increase Economic Welfare? *Russell Cooper* 138

The Economics of Rising Stars *Glenn M. MacDonald* 155

Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach *Assar Lindbeck and Dennis Snower* 167

Bertrand Competition for Inputs and Walrasian Outcomes *Dale O. Stahl II* 189

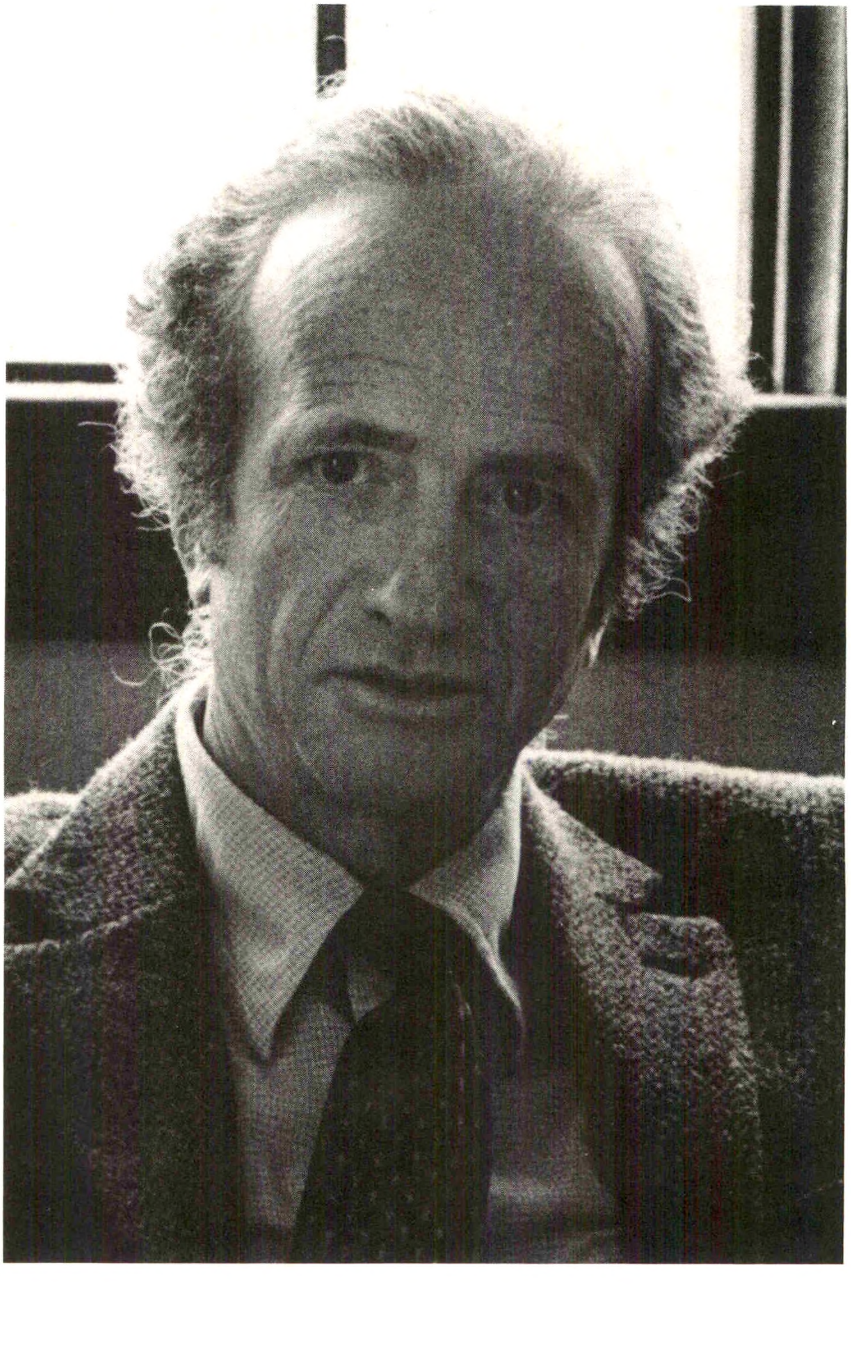
- Submit manuscripts (4 copies), 50 pages maximum, double-spaced, to:
  Orley Ashenfelter, Editor, *AER*; 209 Nassau Street, Princeton, NJ 08542-4607.

- Copyright © American Economic Association 1988. All rights reserved.

- Submission fee: $50 for members; $100 for nonmembers. Please pay with a check or money order payable in United States Dollars. Canadian and Foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Style guides will be provided upon request.

Shorter Papers

On the Interpretation of Near Random-Walk Behavior in GNP *Kenneth D. West* 202
Determinants of the Equilibrium Real Exchange Rate *Peter Neary* 210
Nonadditive Preferences and the Marginal Propensity to Consume *Ehud I. Ronn* 216
Farmers' Cooperatives as Behavior Toward Risk *Miles S. Kimball* 224
Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom? *Uzi Segal* 233
Uncertain Tax Policies, Individual Behavior, and Welfare *James Alm* 237
Tobin's *q* and the Importance of Focus in Firm Performance *Birger Wernerfelt and Cynthia A. Montgomery* 246
Oligopoly and Vertical Integration: Note *Y. Joseph Lin* 251
Families and the Economics of Risks to Life *Maureen L. Cropper and Frances G. Sussman* 255
Some Welfare Implications of Job Mobility in General Equilibrium *Earl L. Grinols and Steven J. Matusz* 261
Is Increased Price Flexibility Stabilizing?
Comment *Stephen R. King* 267
Reply *J. Bradford De Long and Lawrence H. Summers* 273
The Choice Among Medical Insurance Plans:
Comment *Frank D. Gianfrancesco* 277
Reply *Yael Benjamini and Yoav Benjamini* 281
The Disinterest in Deregulation:
Comment *Joe A. Bell* 282
Reply *Robert E. McCormick, William F. Shughart II, and Robert D. Tollison* 284
Price Leadership and Welfare Losses in U.S. Manufacturing:
Comment *Vaughan Dickson* 285
Reply *Micha Gisser* 288
Advertising and Economic Welfare:
Comment *Oded Hochman and Israel Luski* 290
Diamonds Are a Government's Best Friend: Burden-Free Taxes on Goods Valued for Their Values:
Comment *David Friedman* 297
The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67: A Critique and Update of Wolff's Estimates *Fred Moseley* 298
Reply *Edward N. Wolff* 304



Gary S Becker

Family Economics and Macro Behavior[†]

By GARY S. BECKER[*]

It is tempting to use the audience captured by a presidential address to pontificate about the sad state of economics. You probably will conclude that I have surrendered to the temptation. But I do recognize that my good luck in becoming president of our Association does not automatically endow me with commanding wisdom over all of economics. I will do my best to stick to my knitting. And for many years much of my research has been directly toward investment in human capital and the understanding of family behavior.

Modern economists neglected the behavior of families until the 1950s. Since then economic analysis has been used to explain who marries whom and when (if ever) they divorce, the number of children and investments in each child's human capital, the extent and timing of labor force participation by married women, when elderly parents rely on children for support, and many other family choices. A fair conclusion, I believe (need I remind you of my biases?), is that the economic approach contributes important insights toward explaining the large decline in birth rates during the past 100 years, the rapid expansion in the labor force participation of married women after the 1950s, the explosive advance in divorce rates during the past two decades, and other major changes in the family. Family economics is now a respectable and growing field.

Yet perhaps because family economics is a new field, only a small literature considers the implications for other parts of economics. The family is such an important institution that progress in understanding how it behaves is justification enough for any discipline. But most economists, including the audience here, are not particularly concerned about family behavior. Your interest must be stimulated through a demonstration that its study helps in the analysis of other problems.

In this address I try to maintain your interest by exploring the contribution to macroeconomics from the progress in family economics. This is a challenge not only because macro behavior is a central part of economics but also because its link to the family may seem remote and unimportant. By macroeconomics I mean the analysis of economywide behavior. Much of the time is spent on long-term economic growth, although I also discuss short and long cycles in economic activity, and the interaction between overlapping generations through Social Security, transmission of inequality, and in other ways.

Of course, one paper even by a macro expert cannot do justice to these topics, and I do not pretend to be such an expert. My purpose is to help you recognize that many conclusions in these and presumably other macro areas change radically when family choices get the attention they deserve. I apologize for the technical nature of some of the discussion that may seem out of place in a presidential address.

I. The Malthusian and Neoclassical Models

In considering the relation between economic growth and the family, it is natural to begin with Thomas Malthus's great contribution. Although usually called the Malthusian theory of population growth, a more appropriate name is the Malthusian theory of wages and average income. His first monograph, subtitled "With Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers," begins with an objection to the conclusion of these writers that the economic position of mankind will continue to improve over time. In the process of rebutting their arguments, Malthus develops his famous theory of population growth and reaches much more pessimistic conclusions about the long-term economic prospects of the average family.

[†] Presidential address delivered at the one-hundredth meeting of the American Economic Association, December 29, 1987, Chicago, Illinois.

[*] Departments of Economics and Sociology, University of Chicago, Chicago, IL 60637. I received valuable comments from Robert Barro, Edward Prescott, Sherwin Rosen, George Stigler, and Robert Townsend. I was supported by the National Science Foundation under grant no. SES-8520258 and by National Institute for Child Health and Development grant no. SSP 1 R37 HD22054. I received useful research assistance from Michael Gibbs.

You will recall that the Malthusian model assumes diminishing returns to increases in the level of population—that is, to increases in employment—when land and other capital are fixed. The analytical heart of his model (I am not concerned with the details of what he actually said) is consistent with constant returns to the scale of labor and capital, as long as the capital stock, including usable land, does not respond to changes in wages and interest rates.

The response of fertility and mortality to changes in income determine the Malthusian supply of population. Population grows more slowly when wages are low because the average person marries later and thereby has fewer children (the preventive check on population), and also because deaths increase when families are poorer (the positive check). Historical studies indicate that the effect of the economy on age at marriage was considerably greater, at least in Europe, than was its effect on death rates (see Ronald D. Lee, 1987b, pp. 450–51). Therefore, I will ignore the positive effect and consider only the preventive check through changes in the number of children.

The long-run equilibrium wage rate is found at the point on the positively inclined population supply curve where the average family has two children. The economy's production function then determines the stationary level of population that is consistent with this long-run wage rate. There is no presumption that this equilibrium wage is at the subsistence level, especially if the positive check through death rates is not important. In this model tastes for marriage and children, not vague notions of subsistence, determine long-run wages.

The long-run wage is stable in the Malthusian model when shocks push the system out of equilibrium. For example, if an infectious disease destroys much of the population, as the Black Death destroyed perhaps 25 percent of certain European populations during the fourteenth century, the decline in population raises the marginal productivity of labor. The resulting rise in wages encourages families to marry earlier and have more children. Population begins to grow and its increase over time lowers wage rates back toward equilibrium. Ultimately, this dynamic process restores both the wage rate and the level of population to their long-run levels.

If the amount of usable land increases, wages rise and that stimulates higher birth rates. Again, the growth in population continues to lower wage rates until eventually the long-run wage is restored. However, population is permanently higher because the amount of land is greater.

This example brings out that the equilibrium wage is more immune to shocks in the Malthusian system than is the level of population. Indeed, if tastes are stable over time—the Malthusian model, along with George Stigler and myself (1977), assumes *de gustibus non est disputandum*—and if technology does not continue to improve, the equilibrium wage rate remains fixed by the point on the stable supply curve where the typical couple has two surviving children. The Malthusian model does help some in explaining very long-term changes in European wage rates prior to the nineteenth century (Lee, 1987b, gives a good analysis of the evidence). People evidently married earlier when wages were above the equilibrium level and married later when they were below.

It is ironic that Malthus's first essay on population was published in 1798 at the close of the eighteenth century. Although his system was accepted by many leading economists of the nineteenth century (see John Stuart Mill, 1848, Book I, ch. X), events after publication were not kind to the theory. Fertility eventually fell sharply rather than rose as wage rates and per capita incomes continued to advance during much of the nineteenth and twentieth centuries in the United States, Western Europe, and Japan.

The contradiction between the theory and events explains why most economists during the first half of this century showed little interest in explaining long-term trends in income and population. But the subject is too important to remain neglected, and Robert Solow, David Cass, and others developed the neoclassical growth model in the 1950s and early 1960s. This model incorporates two major advances over the Malthusian model. Each person maximizes utility that depends on present and future consumption. More important is the recognition that changes in the capital stock respond to rates of return on investments. Unfortunately, the neoclassical model also takes a sizable step backward from Malthus by assuming that fertility and other dimensions of population growth are independent of wages, incomes, and prices.

I trust that the basic properties of a simple neoclassical model are familiar. What may not be generally appreciated is that despite the different assumptions, the analytic structures of the neoclassical and Malthusian models are quite close and many of their implications are similar. If technology and preferences do not change over time, both models have stable steady-state levels of per capita income. The neoclassical equilibrating mechanism works through changes in the rate of investment, while the Malthusian mechanism works through changes in the rate of population growth. To illustrate, if the capital-labor ratio exceeds its steady-state level, the rate of return on capital is below and the wage rate is above their steady-state levels. In the neoclassical model this discourages investment, which lowers the capital-labor ratio over time (with exogenous population growth). In the Malthusian model this encourages population growth, which also lowers the capital-labor ratio over time (with exogenous investment in capital). We have seen that a shock to population in the Malthusian model has no effect on the level of population or per capita income in the long run. Similarly, in the neoclassical model a shock to the capital stock (perhaps wartime destruction of capital) has no long-run effect on the aggregate capital stock or per capita income.

The persistent growth in per capita incomes during the past two centuries is no easier to explain within the neoclassical framework than within the Malthusian. Of course, the neoclassical model postulates exogenous technological progress to "explain" continuing growth in per capita incomes, but the need to rely on "exogenous" progress is a confession of failure to explain growth within the model. Moreover, the Malthusian model can equally well postulate exogenous progress to "explain" persistent growth.

II. The Family and Economic Growth

After a short while the economics profession became disenchanted with the neoclassical model, presumably because it too did not help in understanding progress. The excitement reflected in hundreds of papers that extended and elaborated this model in the 1950s and 1960s gave way during the past fifteen years to a lack of interest in the analytics of growth that is a little reminiscent of the situation during the first half of the century.

Fortunately, a more relevant growth model is available through combining the best features of the neoclassical and Malthusian models and by adding a focus on investment in knowledge and skills. The neoclassicists are right to emphasize endogenous capital accumulation and utility maximization. Malthusians are right to stress the response of fertility and other components of population growth to changes in the economy, and that these responses can greatly influence economic change.

I will sketch out a modified neoclassical model where parents choose both the number of children and the capital (human or physical) bequeathed to each child. Parental altruism or "love" toward children provides a powerful framework for the analysis of both the quantity and so-called quality of children. Altruism means that the utility of parents depends on the utility of each child. The assumption of altruism is realistic for the vast majority of families, although parent-child interactions are determined also by other motives. Presumably, the altruism per child is negatively related to the number of children, so that an additional child lowers the utility per child to parents in the same way as (please excuse the analogy) an additional car lowers the utility per car.

Such altruism is easily grafted onto the neoclassical utility function by letting parents' utility depend on their own life-cycle consumption and separately on their degree of altruism per child, the number of children, and the utility of each child. This formulation has the important implication that preference for parents' relative to children's consumption (so-called time preference) is not exogenous but rises as the number of children increases.

The resources available to parents from the capital they inherit and labor earnings are spent either on own consumption, on the costs of rearing children, or on transfers to children of human and other capital. Since child rearing is time intensive, the cost of rearing children is positively related to the value of parents' time. Income per capita would rise between the parents' and the child's generations if the total capital bequeathed to each child exceeds the capital inherited by each parent.

Parents choose optimal values of their own consumption, the number of children, and capital transferred to each child while taking into account the cost of rearing children and the dependence of their utility on the utility of children. This analysis has many implications for the behavior of fertility that Robert Barro and I explore elsewhere (see 1987 and 1988). Here I concentrate on a few that alter implications of the neoclassical model about capital accumulation and growth.

If the number of children demanded by the typical family is positively related to the income of parents (the Malthusian assumptions), or at least if it is not strongly negatively related, then this model also has stable steady-state levels of the capital-labor ratio and per capita income. But these steady states depend on variables that change the demand for children.

One example is the consequences of an extended but temporary decline in income and productivity—perhaps due to the disorganization induced by a lengthy depression. In the neoclassical model this has no long-run effect on either per capita or aggregate income. In our modified model an extended decline in productivity can permanently lower *aggregate* income because birth rates may fall when productivity, wages, and interest rates fall. Recall the sharp decline in birth rates during the Great Depression.

Just over a decade age, Barro (1974) showed that a dose of family economics radically alters traditional conclusions about the effects of budget deficits on private savings. For example, deficits to finance Social Security payments tax future generations to support the elderly. Altruistic parents who leave bequests to their children do not seek an intergeneration redistribution of incomes, so they would increase their bequests to offset the effect on children of future taxes. If these families are common, Social Security payments and other public expenditures financed by taxes on future generations would not have much effect on private savings. This is the so-called Ricardian equivalence theorem.

A larger dose of family economics gives more radical implications in some respects but also has more conventional implications for the relation between Social Security and savings. Various comments on Ricardian equivalence emphasize that some families do not leave bequests; I will discuss these families in Section IV. Development economists have long recognized that parents value children who provide support during old age. A Social Security system that replaces child support of parents with public support raises the net cost of children to parents (not to society) since they are no longer as useful to elderly parents. As a result, a Social Security system tends to reduce the demand for children. Social Security also reduces the demand for children by parents who do not receive support but provide bequests. The net cost of children to these parents also increases when they raise bequests to offset the effect of Social Security taxes on children.

For reasons given earlier, a lower demand for children raises the capital bequeathed to each child. Therefore, Social Security and other public transfers between generations would *raise* private savings *per child*, and as a result, raise wage rates and the capital-labor ratio in the next generation. Yet total private savings of the present generations would fall, as in a conventional life-cycle analysis with no bequests, if the decline in fertility exceeds the greater saving for each child.

Consider next an example from tax incidence. A tax on income from capital initially lowers after-tax returns and discourages investment. In the neoclassical model, capital then falls over time until the after-tax rate of return again equals the given rate of time preference. In public finance jargon, a tax on capital would be fully shifted in the long run.

A difficulty with this conclusion is the neoclassical assumption that fertility is fixed, which is especially inappropriate for very long-term changes in incidence. Fertility would fall as capital fell in response to the tax if fertility is positively related to per capita income. A fall in fertility lowers preference for present consumption and raises the demand for investment in each child through the interaction between the quality and quantity of children. Then the equilibrium after-tax rate of return must also fall, and the tax on capital is only partially shifted even in the long run.

The conclusion is more radical if fertility is negatively related to per capita income (for reasons discussed next). Fertility then increases when the stock of capital falls. Since the increase in fertility lowers investment per child, the equilibrium after-tax rate of return would have to increase. That is, we have the paradox that a tax on capital is eventually shifted by more than 100 percent! Let me assure the theorists that this strange result does not violate the second-order conditions.

Does a negative relation between fertility and per capita income imply that children are an "inferior" good (to use the economist's infelicitous language)? The answer is no because the cost of rearing children increases when the capital-labor ratio and per capita income rise since wage rates and the value of parents' time spent on children rise along with the capital-labor ratio. Fertility would fall if the positive effect on fertility of an increase in income is weaker than the negative effect due to the rise in cost. The substitution effect often dominates the income effect in rich countries, for child care in these countries requires considerable time and energy of parents.

If fertility is negatively related to per capita income, an increase in the capital-labor ratio above its steady-state level would reduce fertility and thereby encourage more investment per child. The capital-labor ratio would continue to increase over time if this positive effect on investment dominates the negative effect of a lower rate of return. Consequently, a negative relation between fertility and per capita income can destabilize what is otherwise a stable steady state (see the formal analysis in Robert Tamura, 1986).

Demographers have long been aware that fertility eventually declines as a country develops. Less well appreciated (although see the earlier literature by R. R. Nelson, 1956; Robert M. Solow, 1956, pp. 90–91; S. C. Tsiang, 1964, and others on low-level "traps") is that a negative relation between a country's fertility and its income can destabilize a steady-state equilibrium and cause a protracted period of rising per capita incomes. However, although a decline in fertility is an important stimulus in early stages of development, it alone cannot explain sustained growth over a century or longer. In the absence of other forces, a growing economy with neoclassical production functions but without continuing technological progress eventually moves to a stable steady state with low fertility and high per capita incomes.

A promising approach to sustained growth that complements the role of fertility builds on the special properties of education and other learning. The important property for this purpose is that investments in education and other human capital are more productive when past investments are larger. That is to say, accumulation of knowledge and skills in the past eases the acquisition of additional knowledge. The mastery learning concept in education pedagogy uses this property to organize the teaching of mathematics and other subjects to children (see Benjamin S. Bloom, 1976). Such a production technology implies that rates of return on investments in human capital may not fall and may even rise as the stock of human capital grows.

Perhaps it was reasonable in Malthus's time to neglect investments in human capital, but there is little excuse for the neglect in neoclassical growth theory. Modern economies spend enormous amounts on education and other training of children, and parents' investments in children are a far more important source of an economy's capital stock than are bequests or the life-cycle accumulation of physical capital. Dale Jorgenson and Barbara Fraumeni (1987) estimate that human capital comprises over 70 percent of the total capital stock in the United States. This estimate may be too low because it does not include the contribution of human capital to output in the household sector (the authors do try to estimate household output). Seventy percent may be higher than the true fraction because it makes no allowance for the contribution of "raw labor" to output. I would guess that the true ratio of human capital to the total capital stock may be as high as 90 percent or as low as 50 percent. Of course, even this lower percentage signifies a large contribution. The neglect of human capital in wealth and income accounts greatly distorts comparisons of savings propensities and the accumulation of wealth.

Only recently have growth models begun to appreciate the potential of the learning-by-having property of human capital for generating sustained growth in per capita incomes (see Paul Romer, 1986; Robert E. Lucas, Jr., 1988; and Robert G. King and Sergio Rebelo, 1986; pioneering earlier work includes Kenneth J. Arrow, 1962; Yoram Ben-Porath, 1967; Hirofumi Uzawa, 1965; and Sherwin Rosen, 1976). Kevin M. Murphy and I are developing an analysis that combines such a human capital technology with unskilled labor, physical capital, and endogenous fertility that results from altruism. (See Gary S. Becker, 1971, pp. 204, 207–208, for an earlier effort to combine human capital, unskilled labor, and physical capital.) Our model has a "Malthusian" equilibrium where per capita income is constant and low and fertility is high. However, if this equilibrium receives big enough technology and other shocks—good luck may be required—the economy takes off toward a perpetual growth equilibrium with a decline in fertility and increased investment per child. Knowledge continues to grow through its embodiment in additional human capital.

Family economics is critical to the analysis since choices about number of children and investments in each child's human capital helps determine whether the economy ends up at a "good" (i.e., growth) equilibrium or at a "bad" (i.e., Malthusian) equilibrium. Obviously, we do not have the full answer to economic growth—public policies, conglomeration effects, and other considerations are surely important—but I do believe that our story contributes a sizable part of the answer.

III. Short and Long Cycles

Let me now turn briefly to the relation between family behavior and cycles in aggregate output and other variables. For centuries marriages, births, and other family behavior have been known to respond to fluctuations in aggregate output and prices. In an early use of regression analysis in the social sciences, G. Udny Yule (1906) demonstrated that English marriages and births in the nineteenth century moved together with the business cycle. Subsequent studies showed that higher order as well as first births, divorce rates, and possibly the labor force participation of secondary workers all fluctuated procyclically in many countries (see, for example, Becker, 1960, and Morris Silver, 1965). Birth rates in the United States apparently became countercyclical after many married women entered the labor force. Children are cheaper during recessions because the value of time spent on children by working mothers is low then (see William P. Butz and Michael P. Ward, 1979). Investments in education and other human capital are much less procyclical than investments in physical capital also because the foregone value of time spent in school is cheaper during bad times (Linda N. Edwards, 1975).

Of course, none of the competing macro models of business cycles—be they Keynesian, monetarist, neoclassical, or real—rely on family behavior to cause business cycles. However, declining population growth was a major cause of the secular stagnation feared by Alvin H. Hansen (1939) in his presidential address to our Association almost fifty years ago. Family behavior may play more than a negligible role even in generating ordinary business cycles. For example, an increase in the labor supply of married women or young people when household work or school becomes less attractive can induce cyclical responses in aggregate output and other variables. Cycles started by shifts in labor supply induce a negative relation between wage rates and aggregate output over business cycles. This would help explain why cyclical fluctuations in real wages appear to be less positively related to cyclical fluctuations in aggregate output than is implied by business cycles models that emphasize the demand side.

Although family behavior presumably has only a small part in the generation of ordinary business cycles, it is likely to be crucial to long cycles in economic activity. Malthus claimed that family choices cause long-term fluctuations in the economy through the lagged effects first of marriages on births and then of births on the size of the labor force (see Maw Lin Lee and David Loschky, 1987). Modern demographic analysis generates long cycles in population growth rates through the relation between aggregate fertility and the age distribution, and perhaps also between fertility and the size of a cohort (see, for example, James C. Frauenthal and Kenneth E. Swick, 1983, and Ronald Lee, 1987a). In our modified Malthus-neoclassical model, family choices cause long cycles not only in population growth, but also in capital, output, and other variables if the elasticity of the degree of altruism per child with respect to the number of children declines as families get larger, a reasonable assumption. Fertility and per capita income then fluctuate in generation-long cycles whenever the economy is disturbed away from the steady state (for a proof, see Jess Benhabib and Kazuo Nishimura, 1986).

In the 1920s, the Russian economist Nicholas D. Kondratieff claimed that capitalist economies exhibit long-term fluctuations of about fifty years' duration in output and prices (see Kondratieff, 1935). Simon Kuznets (1958) later argued that long-term fluctuations only last about twenty years. If long cycles of the Kondratieff or Kuznets type exist—we will need another 200 years of data to determine whether they do exist or are just a statistical figment of an overactive imagination—they almost certainly will depend on fertility and other family decisions that biologically require a long time to implement.

IV. Overlapping Generations

The intrinsic risks faced by the elderly, sick, and unemployed are surely no greater in rich countries like Germany and the United States than in poor countries like China and India, nor do these risks rise as a country develops. Yet the first large-scale Social Security program was introduced by Germany a mere 100 years ago. China, India, and numerous other countries still have only modest programs that exclude many of their old people. We take publicly financed schools for granted, but they were unimportant until the latter half of the nineteenth century. Public and private programs that protect against the consequences of illness and unemployment are even newer and less common than Social Security and public schools.

Throughout history the risks faced by the elderly, young, sick, and unemployed have been met primarily by the family, not by state transfers, private charity, or private insurance. Children usually cared for elderly or infirm parents, the unemployed looked to their families for temporary support, and parents have spent much time, money, and energy to rear and train their children. Despite the rapid growth of Social Security payments in the past few decades, almost 20 percent of women aged 65 and over in the United States still live with their children.

The altruism and love of parents, children, spouses, and other relatives have helped protect family members against the hazards of childhood, old age, and other risks. When altruism is insufficient—unfortunately, it often is—what sociologists call social norms frequently emerge that pressure children, parents, spouses, and other relatives into helping out family members in need. In addition, family members use their frequent interaction with one another to raise the level of guilt experienced by a member when he or she does not help out.

The formal analysis of the interaction among overlapping generations began with Paul A. Samuelson's brilliant paper in 1958. This spawned an enormous literature that continues up to the present. Although Samuelson had relevant obiter dicta about social compacts, altruism, and family obligations, his model and that of most of the subsequent literature assumes that each person enters the analysis as a young adult without personal connections to older cohorts. A long review of overlapping generation models in the recent *New Palgrave Dictionary* (see John Geanakoplos, 1987) has no discussion whatsoever of familial relations between members of overlapping generations. I claim that the neglect of childhood and of the intimate relations among parents, children, husbands-wives, and other family members misled these studies sometimes into focusing on minor problems and diverted attention away from some important consequences of the overlapping of generations (the discussion in the next few paragraphs draws partly on Becker and Murphy, 1988).

One example of the emphasis on unimportant problems is the concern with the plight of older people when there are few durable assets that can finance consumption at old age. In an influentual literature on the demand for money, the social role of money is even attributed to a durability that enables older people to finance consumption by selling to the next generation money accumulated when young (see, for example, Thomas Sargent, 1987, ch. 7, and Neil Wallace, 1980). Yet when anthropologists study simple societies that do not have money or other durable assets, they find that old people finance their consumption mainly by relying for support on children and other kin. Indeed, children have been an important resource and money balances an unimportant resource of the elderly in practically all societies, whether simple or complicated.

General equilibrium theorists are concerned about the continuum of equilibria, inefficiency, and other problems that arise in models where overlapping-generations persist indefinitely into the future (see, for example, Geanakoplos, 1987, or Timothy J. Kehoe, 1987). Although these problems would not completely disappear, I conjecture that they would be much less important if overlapping-generations models incorporated the informal trades and assistance available to parents, children, and other members of the same family.

Ever since Plato's *Republic*, philosophers have worried about whether parents invest sufficiently in the health, skills, and morals of their children. Overlapping-generations models usually neglect childhood and concentrate on savings by young adults and their trades with old adults. The treatment of children by parents not only is so important in its own right, but it also greatly influences the relations between older and younger adults (Allan Drazen, 1978, is one of the few earlier studies that recognizes the importance of investments in children for overlapping-generations models.)

I cannot do more on this occasion than present the bare bones of an analysis of how families respond to the demands of both old age and childhood. The analysis is straightforward when altruistic parents leave bequests to their children. The combination of altruism and bequests eliminates any difficulties in financing the wealth-maximizing investment in children's health, training, and other human capital. For if the marginal rate of return on additional human capital exceeds the rate on assets, both parents and children would be better off with additional capital. Parents can save less to offset the negative effect on their consumption of greater spending on their children's human capital, and they can reduce bequests to offset the effect of lower savings on consumption at old age.

Bequests also partly insulate parents from many risks of old age. The opportunity to draw on bequests provides an annuity-like protection against an usually long life and other risks of old age. For example, parents who live longer than expected reduce bequests to help finance consumption in the additional years. If bequests are not a large part of children's assets, bequests can give elderly parents excellent protection against various hazards, and yet changes in bequests do not have much influence on children's welfare. In effect, children help support their parents in old age, although their support is not fully voluntary.

The analysis is less simple when parents do not leave bequests, perhaps because they are not very altruistic or because they expect their children to be better off than they are. These families tend to underinvest in children and underprotect parents against the hazards of old age because bequests are not available to finance investments and old-age support.

Social norms, feelings of guilt, and similar mechanisms may greatly moderate the degree of underinvestments and underprotection. They can induce even selfish parents to invest in children and selfish children to care for sick or poor parents. Economists neglect concepts like norms and guilt because no one really knows how they evolve. Moreover, sociologists (perhaps I should say "we" sociologists since I am now officially also a sociologist) are too prone to use norms as a deus ex machina to explain behavior that is difficult to explain in other ways. Nevertheless, there can be little doubt that norms and other intangible mechanisms do greatly affect the relations between family members in many societies, although presumably, they do not work as well as bequests in linking generations together.

Parents in richer countries have more resources to spend on children and to protect against the hazards of old age. Why then have public expenditures on both the young and old grown rapidly during the 100 years as western countries as they have become richer? One reason is that social norms are weaker in the anonymous urban communities of industrial countries where elderly parents often live far from adult children. A more analytically tractable reason is the high rates of return in modern industrial societies on investments in the health and training of children. Recall my discussion of the role of human capital in economic development. Parents are eager to finance profitable investments in children called for by economic development, as long as they can draw on gifts and bequests that they would give to children. But gifts and bequests would become nil in many families that invest a lot in their children. These families would underinvest in children, particularly when pressure from norms is weak. The growth in public support of schooling and other investments in children as countries develop would then appear to be mainly a response to the positive effect of economic development on the benefits from human capital.

Since families that do not leave bequests are vulnerable to the hazards of old age, it is not difficult to understand why public expenditures on Social Security and medical care for the elderly have also grown rapidly in industrial countries. However, you may be surprised to find out that public expenditures on the old have not been at the expense of the young. Since 1940 in the United States, the ratio of expenditures per child under age 22 to expenditures per adult age 65 or over has hardly changed. Our analysis that combines investments in human capital with old-age support does explain why expenditures on the old and young grew in tandem. By contrast, the popular view of generation fighting—that public expenditures on the elderly grew rapidly because the old became politically powerful as they became more numerous—cannot explain why expenditures on children grew just as rapidly.

The overlapping-generation framework is also a natural one to consider inequality and the transmission of wealth and poverty across generations. Families help perpetuate inequality because children inherit abilities and other "endowments" from parents. Moreover, parents are the major source of the assets and human capital of children. This enormous influence of the family led my esteemed teacher, Frank H. Knight, to claim that "where the family is the social unit, the inheritance of wealth, culture, educational advantages, and economic opportunities tend toward the progressive increase of inequality..." (1935, p. 50).

Abilities and other endowments regress downward from parents to children in successful families where parents earn a lot, and they regress upward in unsuccessful families where parents earn little. The poor underinvest in each child also because they have larger families and less stable marriages. Therefore, children from poorer families tend to earn more than their parents but below the average of their generation, and children from richer families tend to earn less than their parents but above their generation's average.

Earnings depend not only on endowments but also on investments in human capital. Our earlier analysis implies that richer families do not tend to underinvest in their children's human capital because these families leave gifts and bequests. Poorer families do tend to underinvest in children because they are not likely to leave gifts and bequests. The poor underinvest in each child also because they have large families and less stable marriages. Therefore, the relation between the earning of fathers and sons in richer families would depend mainly on the relation between endowments, while the relation between earnings of fathers and sons in poorer families would depend also on the degree of underinvestment in children. Put differently, without offsetting government subsidies to investments in the human capital of poorer children, low earnings would be more persistent across generations than high earnings—the so-called "culture of poverty" across generations would exceed the "culture of privilege."

In every country with data that I have seen—this includes the United States and several European countries (see Table 1 in Becker and Nigel Tomes, 1986), a few Asian countries, and some Latin American countries (James J. Heckman and Joseph V. Hotz, 1986, consider the evidence for Panama)—earnings strongly regress to the mean between fathers and sons. Probably much less than 40 percent of the earnings advantages or disadvantages of fathers pass to sons, and few earnings advantages or disadvantages survive three generations. Evidently, abilities and other endowments that generate earnings are only weakly transmitted from parents to children. This tendency to go from "shirtsleeves to shirtsleeves" in three generations began long before industrialization and government support of education and other human capital. The fourteenth Arab historian and philosopher, Ibn Khaldûn said (I owe this reference to my wife, Guity Nashat), "Prestige is an accident that affects human beings. It comes into being and decays inevitably.... It reaches its end in a single family within four successive generations." (1958, p. 279)... "As a rule, no dynasty lasts beyond the [span] of three generations." (p. 343)

In all these countries, low earnings as well as high earnings are not strongly transmitted from fathers to sons, and Knight's claim about family life causing growing inequality is inconsistent with the evidence. Still, data for both the United States and England do appear to confirm the implication of our theory that low earnings persist more than high earnings across generations (see W. Stanley Siebert, 1987). Of course, *incomes* of the rich regress down more slowly between generations than do their *earnings* because rich children receive gifts and bequests from parents (see Becker and Tomes, 1986, Table 2).

V. Concluding Remarks

I was attracted to the family by its obvious importance in all countries, no matter what the economic system or stage of development. People spend much of their time in a dependency relation—toward parents when children and toward grown children in old age—marriage is a crucial step for most people, children absorb time, energy, and money from their parents, divorce often causes economic hardship and mental depression, and so forth. Economic studies of the family are growing at a steady pace and they are influencing the way other social scientists look this fundamental institution.

The economic analysis of family behavior stimulated the development of techniques and prospectives that already has affected many parts of microeconomics, especially agricultural and labor economics, but also the study of industrial organization and preference theory. For example, the treatment of marriage as a sorting of men and women into small "partnerships" through a reasonably efficient marriage market influenced the analysis of how workers and managers are allocated to different firms. Viewing divorce as a joint decision by husbands and wives based largely on information gathered from living together encouraged some studies of employment separations to blur the analytical distinction between quits and layoffs and to emphasize the information about working conditions and productivity gathered from on-the-job experience.

The message of this address, however, is not the importance of the family per se, even though family welfare is the principal goal of a well-run economic system. Nor that analytical techniques developed to understand family choices are valuable in other parts of economics. The message is that family behavior is active, not passive, and endogenous, not exogenous. Families have large effects on the economy, and evolution of the economy greatly changes the structure and decisions of families. I illustrated how families and the economy interact through a discussion of economic growth and other issues in macroeconomics. A heightened awareness of the interaction between economic change and family choices will hasten the incorporation of family life into the mainstream of economics.

REFERENCES

Arrow, Kenneth, "The Economic Implications of Learning by Doing," *Review of Economic Studies*, June 1962, *29*, 155–73.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November/December 1974, *82*, 1095–117.

______ **and Becker, Gary S.,** "Fertility Choice in a Model of Economic Growth," unpublished paper, Cambridge: Harvard University, 1987.

Becker, Gary S., "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries*, Princeton: Princeton University Press for the National Bureau of Economic Research, 1960.

______, *Economic Theory*, New York: A. Knopf, 1971.

______, *A Treatise on the Family*, Cambridge: Harvard University Press, 1981.

______ **and Barro, Robert J.,** "A Reformulation of the Economic Theory of Fertility," *Quarterly Journal of Economics*, February 1988.

______ **and Murphy, Kevin M.,** "The Family and the State," *Journal of Law and Economics*, forthcoming, April 1988, *31*.

______ **and Tomes, Nigel,** "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, July 1986, *4*, S1–S39.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, August 1967, *4*, 352–65.

Benhabib, Jess and Nishimura, Kazuo, "Endogenous Fluctuations in the Barro-Becker Theory of Fertility," New York University, 1986.

Bloom, Benjamin S., *Human Characteristics and School Learning*, New York: McGraw-Hill, 1976.

Butz, William P. and Ward, Michael P., "The Emergency of Countercyclical U. S. Fertility," *American Economic Review*, June 1979, *69*, 318–28.

Cass, David, "Optimal Growth in an Aggregative Model of Capital Accumulation," *Review of Economic Studies*, July 1965, *32*, 233–40.

Drazen, Allan, "Government Debt, Human Capital, and Bequests in a Life-Cycle Model," *Journal of Political Economy*, June 1978, *86*, 505–16.

Edwards, Linda N., "The Economics of Schooling Decisions: Teenage Enrollment Rates," *Journal of Human Resources*, Spring 1975, *10*, 155–73.

Frauenthal, James C. and Swick, Kenneth E., "Unit Cycle Oscillations of the Human

Population," *Demography*, August 1983, *20*, 385–98.

Geanakopolos, John, "Overlapping Generations Model of General Equilibrium," *The New Palgrave Dictionary of Economics*, Vol. 3, London: Macmillan, 1987.

Hansen, Alvin H., "Economic Progress and Declining Population Growth," *American Economic Review*, March 1939, *29*, 1–15.

Heckman, James J. and Hotz, Joseph V., "The Sources of Inequality for Males in Panama's Labor Market," *Journal of Human Resources*, Fall 1986, *21*, 507–42.

Jorgensen, Dale W. and Fraumeni, Barbara M., "The Accumulation of Human and Non-Human Capital, 1948–84," Cambridge: Harvard University, 1987.

Kehoe, Timothy J., "Intertemporal General Equilibrium Models," Economic Theory Workshop, University of Chicago, January 1987.

Khaldûn, Ibn, *The Muqaddimah*, Vol. 1, Franz Rosenthal, trans., New York: Basic Books, 1958.

King, Robert G. and Rebelo, Sergio, "Business Cycles with Endogeneous Growth," unpublished paper, University of Rochester, 1986.

Knight, Frank H., "The Ethics of Competition," in *The Ethics of Competition and Other Essays*, Chicago: Allen and Unwin, 1935.

Kontratieff, Nicholas D., "The Long Wages in Economic Life," *Review of Economics and Statistics*, May 1935, *17*, 105–15.

Kuznets, Simon, "Long Swings in the Growth of Population and in Related Economic Variables," *Proceedings of the American Philosophical Society*, 1958, *102*, 25–52.

Lee, Maw Lin and Loschky, David, "Malthusian Population Oscillations," *Economic Journal*, September 1987, *97*, 737–39.

Lee, Ronald D., (1987a) "Population Cycles," in *The New Palgrave Dictionary of Economics*, Vol.3, London: Macmillan, 1987.

______, (1987b) "Population Dynamics of Humans and Other Animals," *Demography*, November 1987, *24*, 443–67.

Lucas, Robert E., Jr., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, *19* (forthcoming, 1988).

Malthus, Thomas Robert, *An Essay on Population*, London: J. M. Deut, 1933.

Michael, Robert T., Fuchs, Victor and Scott, Sharon R., "Changes in the Propensity to Live Alone: 1950–1976," *Demography*, February 1980, *17*, 39–56.

Mill, John Stuart, *Principles of Economics*, London, 1848.

Nelson, Richard R., "A Theory of the Low-Level Equilibrium Trap," *American Economic Review*, December 1956, *46*, 894–908.

Romer, Paul, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, February 1986, *94*, 1000–37.

Rosen, Sherwin, "A Theory of Life Earnings," *Journal of Political Economy*, August 1976, *84*, S45–S68.

Samuelson, Paul A., "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, December 1958, *66*, 467–82.

Sargent, Thomas, *Dynamic Macroeconomic Theory*, Cambridge: Harvard University Press, 1987.

Siebert, W. Stanley, "Inequality of Opportunity: An Analysis Based on the Microeconomics of the Family," forthcoming in R. Drago and R. Perlman, eds., *Microeconomic Issues in Labor Economics: New Approaches*, Brighton, U. K.: Wheatsheaf Books, 1987.

Silver, Morris, "Births, Marriages, and Business Cycles in the United States," *Journal of Political Economy*, June 1965, *73*, 237–55.

Solow, Robert M., "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, February 1956, *70*, 65–94.

Stigler, George J. and Becker, Gary S., "De Gustibus Non Est Disputandum," *American Economic Review*, March 1977, *67*, 76–90.

Tamura, Robert, "The Existence of Multiple Steady States in One Sector Growth Models with Intergenerational Altruism," unpublished paper, University of Chicago, 1986.

Tsiang, S. C., "A Model of Economic Growth in Rostovian Stages," *Econometrica*, October 1964, *32*, 619–48.

Uzawa, Hirofumi, "Optimal Technical Change in an Aggregative Model of Economic Growth," *International Economic Review*, January 1965, *6*, 18–31.

Wallace, Neil, "The Overlapping-Generations Model of Fiat Money," in J. H. Kareken and N. Wallace, eds., *Models of Monetary Economics*, Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Yule, G. Udny, "On the Changes in the Marriage- and Birth-Rates in England and Wales During the Past Half Century: With an Inquiry as to Their Probable Causes," *Journal of the Royal Statistical Society*, March 1906, *69*, 88–132.

The Effects of Fiscal Policies When Incomes Are Uncertain: A Contradiction to Ricardian Equivalence

By MARTIN FELDSTEIN*

This paper shows that when earnings are uncertain the substitution of deficit finance for tax finance or the introduction of an unfunded Social Security program will raise consumption even if all bequests reflect intergenerational altruism. Thus, contrary to the theory developed by a number of writers, an operative bequest motive need not imply Ricardian equivalence. Since there is no uncertainty in the present analysis about the date of each individual's death, this conclusion does not depend on imperfections in annuity markets or on the existence of nonlump-sum taxes or other distortions. Rather it follows from the result derived below that, when future earnings are uncertain, bequests are also uncertain and that consumption therefore rises more in response to an increase in current disposable income than to an equal present value increase in the disposable income of the next generation.

It is useful to begin with a summary of the reasoning to be developed in this paper. The starting point of the analysis is the observation that the level of earnings during the "second half" of an individual's working life cannot be accurately predicted during the earlier years. This is particularly important among individuals in managerial, entrepreneurial, and professional occupations who account for a relatively large share of all savings and bequests. Because of this uncertainty, it is optimal for a younger individual to save more than he would if his expected future income were known with certainty. The uncertainty about future income also implies that an individual during the early stage of his life does not know whether he will later want to make a bequest to his children if he can use an annuity to avoid accidental bequests. But even if all bequests are intended and are motivated only by intergenerational altruism, the uncertainty of the individual's future income means that bequests are uncertain.

This uncertainty of future bequests means that an individual will not generally be indifferent between receiving an additional dollar of income when he is young and having his children later receive an amount with a present value of one dollar. Similarly, a one-dollar increase in his current disposable income will increase his current consumption by more than a rise in his children's income with a present value of one dollar. This in turn implies that a tax cut financed by an increase in national debt that will be serviced by future generations will raise current consumption. Similarly, an unfunded Social Security program that promises a net transfer to the current generation from future generations will also raise current consumption.

Before presenting a formal proof of these propositions, I will review the relevant aspects of the current state of the debate about Ricardian equivalence in the context of an economy in which there is no uncertainty about individual incomes. This is done in Section I. Section II then presents a formal model of consumption and bequest decisions of individuals whose earnings during the second half of their working lives are uncertain. Section III uses this analysis to examine the effects of fiscal policies that

---

*Professor of Economics, Harvard University, Cambridge, MA 02138, and President, the National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge MA 02138.

transfer income to the current generation from the next generation. A numerical illustration is presented in Section IV. There is then a brief concluding section.

I. The Ricardian Equivalence Theorem

Although several economists over the years noted the possibility that the aggregate national debt might not be regarded as a net asset because of the implied future debt obligation and therefore that a tax cut might not induce an increase in consumption,[1] it was Robert Barro (1974) who first presented an explicit model in which finite-lived individuals who make bequests to the next generation will completely offset any intergenerational lump-sum transfer imposed by the government. In Barro's analysis, an individual chooses a path of consumption and a bequest to the next generation by maximizing a utility function that has as its arguments the individual's own annual consumption amounts and the utility of the next generation. A current tax cut that is matched by a rise in national debt which is serviced by taxes on future generations does not change the opportunity set of the representative individual. He can maintain his own consumption path and the utility level of the next generation by saving the entire tax cut and bequeathing it (with accumulated interest) to the next generation. This inheritance allows the next generation to maintain its original consumption path and to provide a bequest to its heirs that maintains that generation's utility level. In effect, the process of bequests makes the series of finite-lived individuals act like an infinitely lived individual. With no change in the infinite-horizon budget constraint, there is no reason to change consumption at any date, thus establishing the equivalence of tax finance and debt finance.

One line of objection to this analysis (see, for example, James Tobin, 1980, and Martin Feldstein, 1982) is that an operative bequest motive is relatively rare because individuals believe that the marginal utility of their own retirement consumption exceeds the marginal utility of bequests to their children. Defenders of Ricardian equivalence reply that bequests are in fact relatively common among the upper-income groups that account for such a large share of total wealth accumulation and point to the evidence of Laurence Kotlikoff and Lawrence Summers (1981) that suggests that most existing wealth can be traced to bequests rather than to life-cycle accumulation.

Andrew Abel (1985), and Zvi Eckstein, Martin Eichenbaum, and Dan Peled (1985) showed that the observation of substantial bequests does not imply an operative bequest motive if the age at which death occurs is uncertain and an annuity market does not exist. Moreover, in such an economy Ricardian equivalence will be violated and fiscal policies will affect consumption. However, annuity markets exist and, even with the less than actuarially fair return estimated by Benjamin Friedman and Mark Warshawsky (1985), older egoistic individuals will prefer annuities to accidental bequests.

Although the observation of bequests in an economy with an annuity market may therefore suggest that there is an operative altruistic bequest motive of the type assumed by Barro, other types of bequest motives have been proposed that do not imply Ricardian equivalence. Douglas Bernheim, Andre Schlaefer, and Lawrence Summers (1984) note that bequests may be made for the "strategic" purpose of maintaining the attention if not the actual affection of children and grandchildren. Laurence Kotlikoff and Avia Spivak (1981) suggest that bequests may be the result of an explicit or implicit contract between aged parents and their children in which the parents agree to leave a bequest if they die before a certain age whereas the children agree to provide support if the parents live beyond that age and therefore exhaust their assets. Alternatively, individuals may make bequests because they regard themselves as "stewards" of the funds that they inherited with a moral responsibility to bequeath at least a similar amount to their own children. Each of these models implies that a fiscal transfer from children to parents (i.e., a tax cut or an increase in Social Security retirement benefits) will not be offset by an equivalent increase in bequests.

---

[1] This group includes Don Patinkin (1956), Martin Bailey (1971), and Merton Miller and C. Upton (1974). It is not clear whether David Ricardo actually believed this to be true; see Ricardo (1951) and Gerald O'Driscoll (1977).

Economists will of course differ in the extent to which they accept the strategic bequest, family annuity, or stewardship theories as an explanation of observed bequests. Although I believe that there is probably some truth in each of these explanations, I doubt that they can explain the observed bequests without reference also to intergenerational altruism. The stewardship theory cannot explain the bequests of those who did not receive inheritances. The family annuity theory may be relevant to some moderate-income individuals who are likely to exhaust their assets during retirement but cannot be applied to the wealthy aged whose assets continue to increase as they get older because their spending is less than their income. The strategic bequest theory is more difficult to reject as the primary explanation of observed bequests but is contrary to the persuasive "evidence" of personal introspection as well as to the less reliable assertions of other prospective donors. Moreover, as has been noted by Barro and others, these other bequest motives have ambiguous implications about the direction of the effect of fiscally imposed intergenerational transfers on current consumption.

Robert Barsky et al. (1986) have shown how the existence of nonlump-sum taxes on subsequent risky income can invalidate Ricardian equivalence and cause a positive marginal propensity to consume out of a deficit-financed increase in disposable income. Income taxes on risky income reduce the variance of future net income, providing an otherwise unavailable insurance to individuals that reduces precautionary savings and increases current consumption. Barsky et al. also show that an analogous result holds when individuals live only one period but are uncertain about the income that their heirs will earn. In that case, the nonlump-sum tax on their heirs' income reduces its variance and therefore, by reducing the expected marginal utility of such income to the initial generation, reduces the desired bequest and increases current consumption. Their analysis is thus fundamentally different from that of the current paper because they do not consider the effect of an individual's own income uncertainty on his desired level of bequests. Moreover, the nonlump-sum nature of the taxes that they consider inevitably introduce a nonneutrality.

Abel (1986) shows how a different type of nonlump-sum tax, a progressive tax on bequests or capital, changes the relative cost of current consumption and bequests and thus introduces an incentive to consume more at the present time.

The present paper shows that none of these departures from the original Barro formulation is necessary to demonstrate that Ricardian equivalence is false and that a fiscally mandated intergenerational transfer from the future to the present implies an increase in current consumption. To establish this, I analyze a simple model in which all bequests are caused by intergenerational altruism (i.e, there are no accidental bequests due to an uncertain time of death and the "strategic," family annuity and stewardship motives for bequests are ignored). All taxes and transfers are lump sum. The only difference from the traditional model is that individuals in the first half of their working lives are uncertain about their earnings in the second half.

II. A Life-Cycle Model with Uncertain Earnings

This section extends the traditional life-cycle model with bequests by recognizing the inherent uncertainty of income in later years.[2] Since the purpose of this paper is to demonstrate the possibility of a contradiction to Ricardian equivalence in a model in which all bequests are motivated by explicit intergenerational altruism, the model analyzed here is a very simple one that serves this purpose rather than a more realistic model designed to explore the response of aggregate consumption, capital accumulation, and bequests to variations in the stochastic properties and predictability of lifetime income.

---

[2] Although there have been several analyses of life-cycle models with uncertain income and asset returns (see, for example, Agnar Sandmo, 1970; Jacques Drèze and Franco Modigliani, 1972; and Robert Barsky, Gregory Mankiw, and Stephen Zeldes, 1986), these have not dealt with the relation between income uncertainty and bequests.

Consider therefore a model in which the individual lives two periods. In the first period he works a fixed amount and receives a certain income y_1, which includes any bequest that he receives. In the second period he also works a fixed amount but earns an amount y_2 that cannot be predicted during the first period of life. The second period of life also contains a fixed interval of retirement before death at a known time. Since the amount of work in the first and second periods and the duration of retirement are all fixed, these quantities need not be specified explicitly. Moreover, the assumption of a known date of death is equivalent to assuming the existence of actuarially fair annuities. Finally, there is no need to distinguish between consumption during the working years of the second period and the retirement years because the analysis here focuses on the way that fiscal transfers affect consumption during the first period when subsequent income is unknown.

The individual's utility depends on his consumption during the first and second periods of his life and on the utility of his children. The essential features of the intergenerational bequest model that establishes Ricardian equivalence when income is not stochastic can be captured by assuming that the next generation is the final one: the children of the current generation make no bequests and bear the full burden of any fiscal transfer to the current generation. The utility of the children can therefore be written as a function of their own consumption. In the current context, replacing this specification with an infinite horizon model with each generation linked to the next through the parents' utility function would only complicate the analysis without changing anything essential.

The simplest specification of the stochastic nature of second-period income is that the individual receives a fixed amount Y_2 with probability p and receives zero with probability $1-p$. It will also eliminate unnecessary notation without changing anything fundamental to assume that the interest rate is zero.

In the first period of life, the individual chooses first-period consumption (c_1) to maximize expected utility. In the second period, the individual observes either $y_2=Y_2$ or $y_2=0$ and, conditional on that observation, chooses second-period consumption (c_2) and a nonnegative bequest $(B \geq 0)$ to maximize utility subject to the budget constraint $y_1-c_1+y_2=c_2+B$.

Since this generation's utility is a function of the expected utility of the next generation, some comments about the next generation are in order. In its first period, the next generation receives income z_1 plus the bequest B from this generation. In its second period, the next generation receives income z_2. Since the next generation makes no bequest, its utility is a function of its own path of consumption and its utility can be written as a function of its budget constraint or (discounted) lifetime income: $\Psi(z_1+B+z_2)$.

With these assumptions, the first-period problem of an individual in the current generation is to choose c_1 to maximize $E[u(c_1, c_2, \Psi(z_1+B+z_2))]$, knowing that in the second period he will choose c_2 and B to maximize $u(c_1^0, c_2, \Psi(z_1+B+z_2))$, where c_1^0 is the value of c_1 chosen in period 1. Note that a positive bequest will be chosen at time 2 only if $u_3\Psi' > u_2$ at $B=0$, that is, if the marginal utility of the first dollar of bequest exceeds the marginal utility of an additional dollar of consumption when the bequest level is zero. The interesting case explored below is the one in which this condition holds when $y_2=Y_2$, but does not hold when $y_2=0$, that is, when the bequest is made only when the second-period income exceeds its expected value.

A more general model would recognize that the incomes of the next generation (z_1 and z_2) are both unknown at the time that the member of the current generation chooses his first-period consumption. The individual's optimization problem would incorporate this uncertainty and would specify that the next generation chooses its optimal first-period consumption in the way that maximizes its expected utility. Explicitly incorporating the uncertainty of z_1 and z_2 in this way would substantially complicate the analysis without changing the basic conclusion that individuals who recognize the uncertainty of their own subsequent income will generally act in a non-Ricardian way. The "young" individual's uncertainty about his own future income implies that he is uncertain about whether he will subsequently make a bequest. The uncertain nature of the bequest implies that the individual's current consumption will respond more to a dollar of current disposable income than to an equal-present-value increase in the income of his prospective heirs.

To derive explicit parametric and numerical results to illustrate the effect of income uncertainty, I assume that the utility function is log-linear:

$$E(u) = \ln c_1 + E\left[\ln c_2 + \alpha \ln(z_1 + B + z_2)\right], \tag{1}$$

where α reflects the weight that the current generation assigns to the logarithm of the income of their prospective heirs. To find the value of c_1 that maximizes expected utility, the individual must follow the stochastic dynamic programming principle of solving the second-period problem first and then using the optimal conditional values of c_2 and B to find the optimal value of c_1. From the vantage point of the second period, c_1 is fixed at c_1^0 and c_2, B must be chosen to maximize $\ln c_2 + \alpha \ln(z_1 + B + z_2)$, subject to the budget constraint $y_1 + Y_2 - c_1^0 = c_2 + B$, if $y_2 = Y_2$ or the constraint $y_1 - c_1^0 = c_2 + B$ if $y_2 = 0$.

A positive bequest will be optimal if and only if

$$\frac{1}{y_1 + y_2 - c_1^0} < \frac{\alpha}{z_1 + z_2}, \tag{2}$$

that is, if the marginal utility of c_2 evaluated at $B = 0$ is less than the marginal utility of increased second-generation income, also evaluated at $B = 0$. The only interesting case in the current analysis is the one in which a bequest is optimal when $y_2 = Y_2$ but not optimal when $y_2 = 0$:

$$\frac{1}{y_1 + Y_2 - c_1^0} < \frac{\alpha}{z_1 + z_2} < \frac{1}{y_1 - 0 - c_1^0}. \tag{3}$$

This case will be assumed in the analysis that follows.[3]

Thus $y_2 = 0$ implies $B^* = 0$ and $c_2^* = y_1 - c_1^0$, whereas $y_2 = Y_2$ implies that B^* maximizes $\ln c_2 + \alpha \ln(z_1 + B + z_2)$, subject to the constraint that $c_2 + B = y_1 + Y_2 - c_1^0$. The first-order condition is

$$-\frac{1}{y_1 + Y_2 - c_1^0 - B^*} + \frac{\alpha}{z_1 + z_2 + B^*} = 0 \tag{4}$$

and implies

$$B^* + z_1 + z_2 = \frac{\alpha}{1+\alpha}\left(y_1 + Y_2 + z_1 + z_2 - c_1^0\right) \tag{5}$$

and

$$c_2^* = \frac{1}{1+\alpha}\left(y_1 + Y_2 + z_1 + z_2 - c_1^0\right). \tag{6}$$

Thus when second-period income is high enough to make a positive bequest optimal, the available resources of the two generations are divided in the ratio α to 1 implied by the parameters of the utility function.

These conditional values of B and c_2 can now be substituted into equation (1) with probability weights p and $1 - p$ to derive the optimal value of c_1. Thus

---

[3] The case in which a bequest is always optimal corresponds to the original Barro analysis despite the income uncertainty whereas the case in which a bequest is never optimal is contrary to the observation that individuals do make bequests.

$$(7)\quad E(u) = \ln c_1 - p\left[\ln(1+\alpha)^{-1}(y_1+Y_2+z_1+z_2-c_1) + \alpha\ln(\alpha/1+\alpha)(y_1+Y_2+z_1+z_2-c_1)\right] + (1-p)\left[\ln(y_1-c_1)+\alpha\ln(z_1+z_2)\right].$$

The first-order condition for the optimal value of c_1 is thus

$$(8)\quad \frac{1}{c_1^*} - \frac{p(1+\alpha)}{y_1+Y_2+z_1+z_2-c_1^*} - \frac{1-p}{y_1-c_1^*} = 0,$$

or, equivalently, the quadratic equation

$$(9)\quad (2+p\alpha)c_1^{*2} - \left[(2-p)(y_1+Y_2+z_1+z_2) + (1+p+p\alpha)y_1\right]c_2^* + y_1(y_1+Y_2+z_1+z_2) = 0.$$

Before analyzing the implications of (9) for the effects of fiscal policy, it is useful to derive the optimal consumption and bequests in the same model but without the uncertainty of second-period income. If the individual knows with certainty at the beginning of his life that his second-period income will be pY_2 (i.e., the mean of the uncertain distribution), he will choose c_1, c_2, and B to maximize $\ln c_1 + \ln c_2 + \alpha\ln(z_1+z_2+B)$, subject to the budget constraint $c_1 + c_2 + B = y_1 + pY_2$ and the nonnegativity constraint on bequests ($B \geq 0$). This implies the optimal values

$$(10)\quad c_1^{**} = \frac{y_1+pY_2+z_1+z_2}{2+\alpha},$$

$$(11)\quad c_2^{**} = \frac{y_1+pY_2+z_1+z_2}{2+\alpha},$$

$$(12)\quad B^{**}+z_1+z_2 = \frac{\alpha(y_1+pY_2+z_1+z_2)}{2+\alpha},$$

as long as the implied value of $B \geq 0$. If the desired bequest is negative, the constraint is binding and the optimum consumption is simply $c_1^{**} = c_2^{**} = (y_1+pY_2)/2$.

III. The Effects of Fiscal Policies

We are now ready to analyze how recognizing the uncertainty of second-period income alters the effects of fiscal policy. Consider therefore a tax cut that raises the first-period disposable income of the initial generation (y_1) and increases the national debt that must be repaid by reducing the first-period disposable income of their children (z_1). Since the interest rate is assumed equal to zero, the debt repayment is equal to the initial tax cut: $dy_1 = dz_1$.

Equation (10) shows that in the case of certainty this fiscal policy has no effect on the first-period consumption of the initial generation: $dc_1 = 0$ because c_1 depends only on the combined endowment of both generations ($y_1 + pY_2 + z_1 + z_2$) and that is unaffected by increasing y_1 and decreasing z_1 by equal amounts. This is the fundamental Ricardian equivalence result of Barro.

In contrast, equation (9) shows that when the uncertainty of second-period income is recognized, a fiscal change that raises y_1 and reduces z_1 by an equal amount will not leave c_1 unchanged. More specifically, a tax cut that raises y_1 but leaves $y_1 + z_1$ unchanged will raise first-period consumption. To see this, note that the solution to equation (9) can be written as[4]

$$c_1^* = \frac{Q - [Qa^2 - 4(2+p\alpha)(y_1 + Y_2 + z_1 + z_2)y_1]^{1/2}}{2(2+p\alpha)}, \tag{13}$$

where $Q = (2-p)(y_1 + Y_2 + z_1 + z_2) + (1 + p + p\alpha)y_1 > 0$. It is straightforward to show that, with $y_1 + Y_2 + z_1 + z_2$ constant,

$$\frac{dc_1^*}{dy_1} = \frac{y_1 + Y_2 + z_1 + z_2 - (1 + p + p\alpha)c_1^*}{[Q^2 - 4(2+p\alpha)(y_1 + Y_2 + z_1 + z_2)y_1]^{1/2}}. \tag{14}$$

Since the denominator is positive, $dc_1^*/dy_1 > 0$ if

$$c_1^* < \frac{y_1 + Y_2 + z_1 + z_2}{1 + p + p\alpha}. \tag{15}$$

But equation (10) showed that when there is no uncertainty about second-period income, the optimal c_1 is $c_1^{**} = (y_1 + pY_2 + z_1 + z_2)/(2+\alpha)$, a smaller quantity than the right-hand side of (15) since for any $p < 1$, $2 + \alpha > 1 + p + p\alpha$ and $pY_2 < Y_2$. Moreover, the existence of second-period income uncertainty increases precautionary saving in the first period and therefore implies that $c_1^* < c_1^{**}$. Since $c_1^* < c_1^{**}$ and $c_1^{**} < (y_1 + Y_2 + z_1 + z_2)/(1 + p + p\alpha)$, inequality (15) is satisfied and therefore $dc_1/dy > 0$; a tax cut balanced by a tax increase on the next generation raises current spending.

Before pursuing the formal analysis any further, it is desirable to ask why income uncertainty causes Ricardian equivalence to fail. When second-period income is uncertain, the individual does not know at the time that he chooses c_1 whether he will ultimately want to make a bequest. If he knew with certainty that he was not going to make a bequest, the extra tax borne by the next generation would be irrelevant to him and he could divide his tax cut between his own consumption in the first and second periods. More generally, the individual raises his first-period consumption (although by less than the increase in disposable income), knowing that with probability $1 - p$ he will not want to make a bequest and will raise his second period consumption by the remainder. With probability p the individual will have high income in the second period, will therefore choose to make a bequest, and will use some of his additional first-period saving to make a larger bequest than he would otherwise have made. Nevertheless, the tax cut raises total consumption of the initial generation and reduces total consumption of the next generation even when a bequest is made.

To see this explicitly, note that equation (5) implies that, when a bequest is to be made, an increase in y_1 and an equal decrease in z_1 implies that the next generation's consumption is reduced by a fraction of the induced consumption

$$\left.\frac{d(B^* + z_1 + z_2)}{dy_1}\right|_{B^* > 0} = -\frac{\alpha}{1+\alpha}\frac{dc_1^*}{dy_1}. \tag{16}$$

Since there is no offsetting change in bequest when $y_2 = 0$ and $B^* = 0$, the average change in second-generation income in response to a current tax cut financed by a tax increase on the second generation is

$$\frac{d(B^* + z_1 + z_2)}{dy_1} = -\left[p\frac{\alpha}{1+\alpha}\frac{dc_1}{dy_1} + (1-p)\right]. \tag{17}$$

The analysis of Social Security retirement benefits is essentially identical in the current context to the analysis of the tax cut. Consider a program that pays a sure benefit β to the current generation in its second period and finances this by a tax of β on the income

---

[4] This is the only feasible solution of the quadratic equation; adding instead of subtracting the square root expression implies a value of c_1 greater than initial income.

z_1 of the next generation. When there is no income uncertainty, the Social Security program raises second-period income to $pY_2+\beta$ and reduces the next generation's initial earnings to $z_1-\beta$, leaving c_1^{**} in equation (10) unchanged. In the case of uncertain second-period income, the payment of a sure second-period benefit is analytically identical to a tax cut. To see this, note that $y_2=0$ now implies $c_2^*=y_1-c_1^0+\beta$ so that the payment of the second-period benefit is the same as an increase in first-period income if it induces the same c_1^0. Similarly, when $y_2=Y_2$, the individual maximizes $\ln c_2+\alpha\ln(z_1-\beta+B+z_2)$, subject to the constraint that $c_2+B=y_1+Y_2+\beta-c_1^0$; this is also identical to the effect of a tax cut that increases y_1 and decreases z_1 as long as it yields the same c_1^0. To see that the optimal first-period consumption is indeed the same, note that the expression to be maximized in equation (7) is modified in exactly the same way by the Social Security program as it would be by a tax-induced rise in y_1 and reduction in z_1. Thus a Social Security program has the same effect of increasing first-generation consumption and reducing the second generation's subsequent consumption as an equal-sized intergenerational transfer achieved by a tax cut.[5]

IV. A Numerical Illustration

A numerical example will illustrate the potential effect of income uncertainty on consumption, on bequests, and on the impact of fiscal policy. The specific example is obviously arbitrary but indicates the potential importance of income uncertainty. In the example, the marginal propensity to consume out of the tax-induced increase in disposable income is almost as large as the average propensity to consume.

Since the results are essentially independent of the units of measurement, I set first-period income equal to unity: $y_1=1$. With $p=0.5$ and $Y_2=2$, the expected value of second-period income is also one. The economic specification is completed by setting the next generation's first-period income and second-period certainty equivalence income both equal to one as well: $z_1=z_2=1$. The analysis will be done for two alternative values of the intergenerational altruism parameters: $\alpha=1$ and $\alpha=3$.

Consider first the case in which there is no uncertainty. Second-period income is $pY_2=1$ and is known with certainty. From equation (12), the optimal bequest is $B^{**}=\max[4\alpha/(2+\alpha)-2,0]$. Thus with $\alpha=1$ the individual gives too little weight to the next generation to make any bequest and $B^{**}=0$. In this case the individual consumes all of his income in each period: $c_1^{**}=c_2^{**}=1.0$. With $\alpha=3$, there is enough weight on the next generation's welfare to induce a bequest: $B^{**}=0.4$ and $c_1^{**}=c_2^{**}=0.8$.

When second-period income is uncertain, the optimal value of c_1 is given by equation (13). With $\alpha=1$, $c_1^*=0.6311$, whereas $\alpha=3$ implies $c_1^*=0.5937$. In both cases, first-period consumption is substantially less than it would be if the same expected second-period income could be anticipated with certainty. This reflects both the precautionary demand for saving (against the risk that $y_2=0$) and the saving for subsequent bequests (if $y_2=2$).

If $y_2=0$, the individual will choose to make no bequest with $\alpha=1$ or with $\alpha=3$. In contrast, if $y_2=2$ the individual will choose a bequest of $B^*=0.1844$ with $\alpha=1$ and $B^*=1.3047$ with $\alpha=3$. Thus with $\alpha=1$ the income uncertainty increases the average bequest from $B^{**}=0$ to $pB^*=0.0922$ and with $\alpha=3$ the income uncertainty increases the average bequest from $B^{**}=0.40$ to $pB^*=0.6524$.

Consider now the effect of a fiscal policy that increases the initial generation's first-period disposable income from $y_1=1.0$ to $y_1=1.1$ and reduces the corresponding disposable income of the next generation from $z_1=1.0$ to $z_1=.9$. This raises the first-period consumption from $c_1^*=0.6311$ with $\alpha=1$ and $y_1=1.0$ to $c_1^*=0.6896$ with $\alpha=1$ and $y_1=1.1$; the increase of 0.0585 implies a marginal propensity to consume out of the fiscal transfer of 0.585, almost as high as the initial average propensity to consume of 0.631. Similarly, with $\alpha = 3$, first-period consumption rises from $c_1^* = 0.5937$ to $c_1^* = 0.6434$, implying a marginal propensity to consume out of the fiscal transfer 0.497, approximately 85 percent of the average propensity to consume.

---

[5] This assumes that the size of the second-period benefit is not so large that the individual wants to consume more than his entire first-period income. In this case, it is still true that the benefit increases first-period consumption but by less than the rise that would result from an equally large tax reduction.

The fiscal transfer induces an increased bequest, although not a large enough increase to maintain the consumption of the next generation. This is true even if attention is limited to the case in which $y_2 = 2$ so that a bequest is made. For example, with $\alpha = 1$ the bequest when $y_2 = 2$ rises from $B^* = 0.1844$ to $B^* = 0.2552$ but the increased bequest of 0.0708 is less than the increased tax of 0.10 paid by the next generation. Moreover, the fiscal transfer only raises the average bequest from $0.5(0.1844) = 0.0922$ to $0.5(0.2552) = 0.1276$, an increase of 0.0354 in comparison to the universal tax rise of 0.1000. Similarly, with $\alpha = 3$, the bequest when $y_2 = 2$ rises from $B^* = 1.3047$ with $y_1 = 1.0$ to $B^* = 1.3674$ when $y_1 = 1.1$. The increased bequest (0.0627) is slightly smaller then with $\alpha = 1$ and offsets less than two-thirds of the tax increase even among those who receive a bequest. More generally, the average bequest rises from $0.5(1.3047) = 0.6524$ to $0.5(1.3674) = 0.6837$, a rise of 0.0313 or less than one-third of the average tax increase.

V. Concluding Comment

This paper has shown how the inability of individuals to predict accurately their subsequent earnings can cause fiscal transfer from future generations to the current generation to raise current consumption. When earnings uncertainty makes individuals uncertain about whether they will subsequently wish to make a bequest, Ricardian equivalence fails. Even an economy in which all bequests are altruistic will behave in a non-Ricardian way when income uncertainty causes bequests to be uncertain.

Unlike the uncertainty that arises because the time of death is unknown, the unpredictability of individual future income cannot be avoided even in principle by an annuity market or other insurance market. The uncertainty of personal earnings is unavoidable because of the moral hazard problem involved in trying to ensure individual earnings.

The very simple model developed in this paper can demonstrate the potential importance of earnings uncertainty and the general inapplicability of Ricardian equivalence. It also shows how the individual's uncertainty about his future income can reduce consumption and increase the probability and expected size of bequests. It would be desirable to extend this analysis to a more realistic specification of uncertainty and to analyze the implications for capital accumulation, income distribution, and fiscal policy in an infinite horizon model.

REFERENCES

Abel, Andrew B., "Precautionary Saving and Accidental Bequests," *American Economic Review*, September 1985, *75*, 777–91.

_______, "The Failure of Ricardian Equivalence Under Progressive Wealth Taxation," NBER Working Paper No. 1983, Cambridge, MA: National Bureau of Economic Research, July 1986.

Bailey, M. J., *National Income and the Price Level*, New York: McGraw-Hill, 1971.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November-December 1974, *82*, 1095–117.

Barsky, Robert J., Mankiw, Gregory and Zeldes, Stephen, "Ricardian Consumers with Keynesian Propensities," *American Economic Review*, September 1986, *76*, 676–91.

Bernheim, Douglas, Schlaefer, Andre and Summers, Lawrence, "Bequests as a Means of Payment," NBER Working Paper No. 1303, Cambridge, MA: National Bureau of Economic Research, March 1984.

Drèze, Jacques and Modigliani, Franco, "Consumption Decisions Under Uncertainty," *Journal of Economic Theory*, December 1972, *5*, 308–25.

Eckstein, Zvi, Eichenbaum, Martin S. and Peled, Dan, "The Distribution of Wealth and

Welfare in a Model with Incomplete Annuity Markets," *Quarterly Journal of Economics*, August 1985, *100*, 789–806.

Feldstein, Martin, "Government Deficits and Aggregate Demand," *Journal of Monetary Economics*, January 1982, *9*, 1–20.

Friedman, Benjamin, and Warshawsky, Mark, "The Cost of Annuities: Implications for Saving Behavior," NBER Working Paper No. 1682, Cambridge, MA: NBER, August 1985.

Kotlikoff, Laurence and Spivak, Avia, "The Family as an Incomplete Annuities Market," *Journal of Political Economy*, April 1981, *89*, 372–91.

______ **and Summers, Lawrence,** "The Importance of Intergenerational Transfers in Aggregate Capital Accumulation," *Journal of Political Economy*, August 1981, *89*, 706–32.

Leland, Hayne E., "Saving and Uncertainty: The Precautionary Demand for Saving," *Quarterly Journal of Economics*, August 1968, *82*, 465–73.

Miller, M. and Upton, C., *Macroeconomics: A Neo-classical Introduction*, Homewood, IL: Irwin, 1974.

O'Driscoll, Gerald P., Jr., "The Ricardian Nonequivalence Theorem," *Journal of Political Economy*, February 1977, *85*, 207–10.

Patinkin, Don, *Money, Interest and Prices*, Evanston, IL: Rowe Peterson, 1956.

Ricardo, D., *The Works and Correspondence of David Ricardo*, P. Sraffa, ed., Cambridge: Cambridge University Press, 1951.

Sandmo, Agnar, "The Effect of Uncertainty on Saving Decisions," *Review of Economic Studies*, July 1970, *37*, 353–60.

Tobin, James, *Asset Accumulation and Economic Activity*, Chicago: University of Chicago Press, 1980.

Asset Prices, Commodity Prices, and Money: A General Equilibrium, Rational Expectations Model

By GLENN W. BOYLE AND LESLIE YOUNG*

An expected utility-maximizing investor spends his portfolio income on commodities and real balances. Commodity prices and asset payoffs are determined endogenously in general equilibrium. The impact of commodity prices on investor welfare yields surprising relationships among the expected returns required on various assets, for example, real (monetary) disturbances can generate a negative (positive) correlation between inflation and equity payoffs but the expected nominal return on the equity can still be less (greater) than the nominal interest rate.

This paper develops a consumption-based model of asset pricing which integrates the real financial and monetary sectors of the economy. Unlike most earlier consumption-based models which treated the probability distributions of asset payoffs and future commodity prices as exogenous,[1] we derive these distributions endogenously within a general equilibrium model by assuming rational expectations. This yields new insights into the implications of inflation and real sector activity for asset pricing.

The first insights concern the role of equities as hedges against inflation. Zvi Bodie (1976, p. 459) defined the effectiveness of common stocks as an inflationary hedge as "the extent to which they can be used to reduce the risk of an investor's real return which stems from uncertainty about the future level of prices of consumption goods." Many other definitions of an inflationary hedge appear in the literature, all emphasizing the ability of the asset to insulate the consumer against price fluctuations by yielding nominal returns which are positively correlated with such fluctuations. Equities have traditionally been regarded as inflationary hedges because their returns were presumed to have this property. Hence the challenge posed by the recent empirical evidence which indicates that this correlation is *negative*.[2]

We shall argue that such a negative correlation can be reconciled with our intuition that equities are desirable elements of a portfolio in an inflationary environment. The empirical evidence of such a correlation appears paradoxical only because the above definitions of an inflationary hedge bypass the central issue in valuing assets in an inflationary environment. An investor is generally not directly interested in reducing uncertainty about real returns as such, so the fact that an asset offers stable real returns need not imply that it commands a premium in financial markets. Rather, an investor chooses his portfolio to maximize his expected utility. The first-order conditions for this maximization imply that an equity will be an attractive element of a portfolio in an uncertain inflationary environment, commanding a price premium over its discounted expected return, if and only if its nominal payoffs are positively correlated with the investor's *marginal utility of nominal income*. The impact of inflation on the latter term is complex, depending on the nature of the shocks generating the inflation and on the preferences of the investor. If there were a negative correlation between the marginal utility of income and inflation, then it would be precisely *because* equities have nominal returns which are negatively correlated with inflation that they would be desirable as hedges against inflationary shocks, commanding a price premium over their discounted expected return, or equivalently, having an equilibrium expected rate of return lower than the nominal interest rate.

---

*Graduate student and V. F. Neuhaus Professor of Finance, Finance Department, University of Texas at Austin. The University of Texas provided partial research support to Glenn Boyle via an Abell-Hanger Endowed Presidential Scholarship and to Leslie Young via a Summer Research Award from the Graduate School of Business. We are grateful to a referee for helpful comments.

[1] Our departures from the literature are discussed in Section I.

[2] Jeffrey Jaffe and Gershon Mandelker, 1976; Charles Nelson, 1976; Eugene Fama and William Schwert, 1977; Zvi Bodie, 1976; and Eugene Fama, 1981, also document a negative correlation between inflation and real stock returns.

The difficulty with assessing the above argument is that the postulated correlations concern the marginal utility of income, which is unobservable. In this paper, we derive the correlation structure of asset payoffs and commodity prices within a general equilibrium model of the underlying economy with explicitly modeled disturbances in output and the money supply. We then determine whether there are ranges of values of the structural parameters of the economy (risk attitudes and income and price elasticities of demand) that yield the correlations postulated. Under real disturbances, there indeed exist parameter values which imply not only a negative correlation between inflation and nominal equity returns (consistent with the empirical results noted above) but also a positive correlation between asset payoffs and the investor's marginal utility of nominal income. Thus, under real disturbances, the observed negative correlation between inflation and nominal equity payoffs is indeed consistent with equities commanding a premium as a hedge against inflationary shocks. On the other hand, if the inflation were due to *monetary* disturbances, then a *positive* correlation between nominal equity payoffs and commodity prices obtains —but equities then, ironically, turn out to be *poor* inflationary hedges, requiring an expected nominal rate of return *higher* than the nominal interest rate because they yield high nominal returns when the marginal utility of nominal income is low.

Our model addresses another important issue in asset pricing: the implications of random disturbances for the Fisher hypothesis that inflation is fully incorporated into nominal interest rates, that is, that the rates of return on real and nominal bonds are equal. Since real bonds have a high nominal payoff when inflation is high, it might appear that the expected nominal rate of return on real bonds required by a risk-averse investor would be lower than that on nominal bonds. However, the critical condition for the latter conclusion to hold is that the nominal payoff on the real bond be positively correlated with the marginal utility of nominal income. We identify ranges of values for the structural parameters of our model where this requirement is violated, so that the expected nominal rate of return required on a real bond is *higher* than that on a nominal bond, both in the case of monetary disturbances and in the case of real disturbances when risk aversion is low. The reverse conclusion in the case of real disturbances holds only when the direct impact of commodity prices on the marginal utility of income is sufficiently weak.

We also address the relationship between the *real* expected rates of return on financial assets. Risk-averse investors will require a higher expected real rate of return on assets which have a relatively high real payoff when the marginal utility of real income is low. Under output disturbances, the latter occurs in a high output state. In such a state, the price level is low so the real payoff on a nominal bond is high. Since the real payoff on a real bond is state-independent, the equilibrium expected real rate of return on a nominal bond exceeds that on a real bond. However, we shall see that high output states could result in low payoffs to equities, so their equilibrium expected real rates of return could be *less* than that on a real bond. Under monetary disturbances, the marginal utility of real income is constant, so all three assets have the same expected real rate of return in equilibrium.

After relating our work to the literature in Section I, we set out the model in Section II and confirm that the impact of risk aversion on asset prices in a consumption-based model depends on the covariance between asset payoffs and the consumer's marginal utility of income. The remaining sections relate this covariance to the structure of the economy. Duality theory provides a neat way of analyzing the equilibrium of the commodity economy and of incorporating the restrictions on consumer demand implicit in our use of real balances in the consumer's utility function. Section III isolates monetary effects by treating a one-commodity monetary model. This analysis is extended to the many-commodity model of Section IV by analyzing consumer choice as a two-stage problem in which the consumer first divides income between goods and money, then allocates goods expenditure across particular goods. Section V introduces monetary disturbances and compares the expected rates of return on equities and on nominal bonds under such disturbances. Section VI considers the pricing of real bonds and compares their equilibrium expected rate of return with that on nominal bonds and on equities. Section VII highlights the importance of modeling money explicitly. Section VIII briefly considers extensions to the model.

I. The Literature

The consumption-based asset pricing models of Robert Merton, 1973; Mark Rubinstein, 1976; Robert Lucas, 1978; and John Cox, Jonathan Ingersoll, and Stephen Ross, 1985, dealt with a one-good economy where commodity prices did not signal relative values and hence had no effect on asset pricing. These models focused on the effects of stochastic investment opportunities on asset pricing, bypassing the effects of the stochastic consumption opportunities associated with fluctuating commodity prices. The latter effects were treated in a many-commodity model by Frederick Grauer and Robert Litzenberger (1979) who obtained valuation formulas for nominal bonds and commodity futures contracts in terms of the covariance between asset payoffs and the marginal utility of income. Since the latter is not directly observable, the empirical implications of such formulas are unclear. Douglas Breeden (1979) extended the two-period Grauer-Litzenberger model to continuous time and, in his multi-good framework, related asset risk premia to the covariance between real asset payoffs and real consumption.[3] As in Grauer-Litzenberger (1979), the probability distributions of the commodity prices were exogenously given: no attempt was made to derive endogenously the stochastic relationship between asset payoffs and commodity prices which determines the key covariances. Thus, Breeden's model recognized (a) the impact of consumption preferences on asset demands and asset prices and (b) the impact of relative goods prices on consumption decisions but did not close the circle by treating (c) the impact of consumption decisions *on* commodity prices and asset payoffs. While showing the possible effects of random commodity prices on asset pricing, he left open the question of whether these effects are consistent with the postulated consumption preferences in general equilibrium. Similar comments apply to the discrete-time model of John Long (1974).

In this paper, we maintain the Grauer-Litzenberger two-period framework, thereby focusing on the implications of consumption risk and variable commodity prices for asset pricing in an exchange economy. In contrast to their model and those of the other authors cited, our model is closed by determining commodity prices and asset payoffs *endogenously* as general equilibrium responses to disturbances in output or the money supply. This permits the key covariances to be analyzed in terms of such disturbances and to be related ultimately to observable structural parameters of the underlying economy. Our explicit treatment of money also represents a significant departure from the above authors, who expressed prices and asset payoffs in "units of account" or "dollar values," even though money played no economic role within their models. Perhaps in recognition of this, they frequently stated their results on asset returns in "real" terms. The empirical interpretation of such results poses the usual problems of choosing a price index. In any case, there remains the empirically important question of how the *monetary* returns to various assets are interrelated. To treat this question in a model where prices play an economic role, we introduce money explicitly into our model of consumer choice and express commodity prices and asset payoffs in terms of money. Real disturbances then have monetary consequences via changes in the price level and the demand for money. This linkage means that real disturbances have stronger effects on prices than in a model which does not treat money explicitly and reversals of the usual implications of risk aversion can occur under less extreme conditions. We shall see that the abstraction from monetary effects would even distort comparisons of the *real* returns on assets. Thus, *the omission of monetary effects in models of asset pricing can be seriously misleading when commodity prices play an economic role.*

---

[3] Douglas Breeden (1979, Sec. 7). Breeden (1986) treats a somewhat different set of issues: the relationship between real and nominal interest rates and the optimal growth paths of aggregate consumption and production.

In a multi-period monetary model with one commodity and unit velocity of circulation of money, Theodore Day (1984) showed that the real payoff on an equity will be negatively correlated with the rate of inflation under real disturbances. Jean-Pierre Danthine and John Donaldson (1986) reached a similar conclusion in a one-commodity model in which real balances enter the investor's utility function (thereby admitting a nonconstant velocity of circulation). Our analysis will determine this correlation within a model which admits a range of commodities and equities.[4] Unlike the last two sets of authors, however, we focus on the implications of this correlation for the consumption-risk hedging properties of various assets; in particular, for the relationship between their equilibrium expected rates of return.

Monetary models in which prices and asset payoffs are endogenously determined have also been developed by Robert Lucas, 1982; Lars Svensson, 1985; and Stephen LeRoy, 1984a,b. We go beyond these authors in providing detailed comparisons of the equilibrium expected rates of return on various assets in terms of the structural parameters of the economy.[5] Our results here complement the work of Lars Hansen and Kenneth Singleton (1982), who proceed in the opposite direction, estimating values for preference parameters from restrictions on the form of the utility function and data on stock market returns.

II. The Model

There are two periods, corresponding to the investment and the consumption decisions of the representative investor. In the investment period, the investor allocates his wealth W between a riskless nominal bond and equities $j=1,\ldots,n$, which constitute titles to the firms producing the goods j. In this period, the input choices of these firms are given but their outputs are uncertain. There are z^j units of equity j, each conferring ownership of θ^j units of good j, where θ^j is a random variable. The following consumption period reveals the value of θ^j and hence the supply $\theta^j z^j$ of good j, after which the investor chooses his consumption. The usual forces of demand and supply determine the prices $p^j(\tilde{\theta})$ of the commodities j and the payoffs $p^j(\tilde{\theta})\theta^j$ on the equities j as functions of the random vector $\tilde{\theta} \equiv (\theta^1,\ldots,\theta^n)$. Our analysis initially focuses on the relationship between the equilibrium expected rate of return on an equity j and the nominal interest rate i, or equivalently, between the price q^j of equity j and its expected riskless present value $E[p^j(\tilde{\theta})\theta^j]/(1+i)$.

---

[4] See fn. 7 and our discussion after Proposition 6′ for the contrast between our results and those of the authors just cited.

[5] Lucas (1982) and Svensson (1985) introduced money into asset pricing via a cash-in-advance constraint. LeRoy (1984a, b) introduced real balances into the utility function. We follow the latter procedure but do not impose LeRoy's assumption that the utility function is additively separable in goods and real balances. Such an assumption is difficult to reconcile with the usual rationalization for real balances in the utility function: that they economize on transactions costs. We shall consider the work of these authors in more detail in the course of stating our own results.

To incorporate money, we suppose that during the consumption period, the representative investor derives a benefit from his holdings of real balances B—defined as his holdings of nominal money M divided by the price level π, which is a function of the vector of commodity prices $p \equiv (p^1,\dots,p^n)$,

$$B = M/\pi.$$

If $X \equiv (X^1,\dots,X^n)$ is his goods consumption vector, then his consumption choices are governed by the utility function

$$U = U(X, B). \tag{1}$$

This utility function can be viewed as the reduced form of a system in which the investor's underlying utility function depends on goods and leisure while, for given goods consumption, leisure increases with his holdings of real balances because they permit him to economize on the time spent in transactions.[6]

If the investor has income I to spend on goods or money, then his budget constraint is

$$\Sigma p^k X^k + \pi(p)B = I.$$

The maximization of the utility function (1) subject to the constraint yields the vector $X(p, I)$ of goods demand functions and the demand function $B(p, I)$ for real balances. Substituting these into the utility function U yields the indirect utility function

$$V(p, I) \equiv U(X(p, I), B(p, I)).$$

This indirect utility function will be used in the investment stage of the decision problem, which we now describe.

If the investor owns Z^j units of equity j and N dollars of nominal bonds, then his portfolio income in state $\tilde{\theta}$ is $\Sigma p^k(\tilde{\theta})\theta^k Z^k + N(1+i)$, where i is the nominal interest rate. During the consumption period, the government introduces the money stock μ into private hands and this represents additional purchasing power for the investor. The investor chooses the Z^j and N to maximize his expected utility subject to his wealth constraint

$$\max_{Z^k, N} E\left[V\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k Z^k + N(1+i) + \mu\right)\right]$$

$$\text{subject to} \qquad \Sigma q^k Z^k + N = W,$$

where q^k is the nominal price of equity k. The first-order conditions determining asset demands are

$$E\left[p^j(\tilde{\theta})\theta^j V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k Z^k + N(1+i) + \mu\right)\right] = \lambda q^j, \quad j = 1,\dots,n \tag{2}$$

$$E\left[V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k Z^k + N(1+i) + \mu\right)\right](1+i) = \lambda, \tag{3}$$

where λ is the Lagrange multiplier associated with the wealth constraint and subscripts denote partial differentiation with respect to the corresponding variable. To simplify our notation, we assume that the outside supply of nominal bonds is zero, although this is not necessary for our results. The supply of equity j is z^j. Since the returns on equities and bonds must equate asset demands to these exogenously given supplies, we can replace N by 0 and Z^j by z^j in the first-order conditions (2) and (3). Eliminating the Lagrange multiplier λ, we then obtain

$$q^j = \frac{E\left[p^j(\tilde{\theta})\theta^j V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu\right)\right]}{E\left[V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu\right)\right](1+i)}. \tag{4}$$

Let cov$[\cdot,\cdot]$ denote the *normalized* covariance between the two arguments, that is, the covariance divided by the mean of each argument. Then (4) becomes

---

[6] For further discussion of the incorporation of money into consumer choices in this way, see, for example, Stephen Turnovsky and William Brock, 1980; LeRoy 1984a; Paul Samuelson and Ryuzo Sato, 1984; and Danthine and Donaldson, 1986.

$$(5)\quad \frac{q^j - E\left[p^j(\tilde{\theta})\theta^j\right]/(1+i)}{E\left[p^j(\tilde{\theta})\theta^j\right]/(1+i)} = \operatorname{cov}\left[p^j(\tilde{\theta})\theta^j, V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu\right)\right].$$

Define the expected nominal (rate of) return e^j required on equity j by

$$1+e^j = E\left[p^j(\tilde{\theta})\theta^j\right]/q^j.$$

By (5)

$$(6)\quad (i-e^j)/(1+e^j) = \operatorname{cov}\left[p^j(\tilde{\theta})\theta^j, V_I\left(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu\right)\right].$$

Equation (5) asserts that the price of equity j exceeds the riskless present value of its expected payoff by a percentage equal to the normalized covariance between that payoff and the marginal utility of income. Equivalently, (6) asserts that the gross nominal return on a nominal bond exceeds the gross expected nominal return required on equity j by a percentage equal to this covariance. Thus, the expected nominal return required on equity j is less than the nominal interest rate if equity j yields a high (low) payoff in states where each dollar has a high (low) marginal value. If the reverse were true, then the expected nominal return required on the equity would be greater than the nominal interest rate.

This conclusion is essentially the same as that of Grauer-Litzenberger (1979) and holds whatever the disturbances generating the probability distributions of asset payoffs and the marginal utility of income. However, by explicitly modeling the relationship between these disturbances and commodity prices and asset payoffs within a general equilibrium model of the economy, we shall be able to identify the structural parameters of the economy which determine the sign of the key covariances. This is facilitated by Lemma 1, which is proved in the Appendix:

LEMMA 1: *Suppose that* $\phi(\tilde{\theta})$ *and* $\Psi(\tilde{\theta})$ *are functions of the vector* $\tilde{\theta}$ *of the independent random variables* θ^j. *If, for each* j, *either both* ϕ *and* Ψ *are increasing functions of* θ^j *or both are decreasing functions of* θ^j, *then* $\operatorname{cov}[\phi(\tilde{\theta}), \Psi(\tilde{\theta})] > 0$. *If for each* j, *one of these functions is increasing and the other is decreasing in* θ^j, *then* $\operatorname{cov}[\phi(\tilde{\theta}), \Psi(\tilde{\theta})] < 0$.

Thus, the sign of $\operatorname{cov}[p^j\theta^j, V_I]$—and hence the relationship between the expected nominal return required on equity j and the nominal interest rate—depends on the signs of the derivatives of $p^j(\tilde{\theta})\theta^j$ and $V_I(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu)$ with respect to the output disturbances θ^j. We shall assume that the investor has rational expectations about the distribution of commodity prices so that the above derivatives can be calculated by differentiating the equilibrium conditions in the commodity and money markets. This calculation would be cumbersome if the equilibrium conditions were stated in terms of ordinary demand functions because both commodity prices and production disturbances also affect demand via the nominal income of the consumer. Instead, we state the equilibrium conditions in terms of the investor's *compensated* demand functions, replacing the nominal income argument of the ordinary demand functions by the investor's utility level. Similarly, we calculate $dV_I/d\theta^j$ by expressing V_I in terms of dual functions, again bypassing the complex effects of θ^j on nominal income. Since the price index used to define real balances can be related to the expenditure function of the consumer (whose derivatives give the compensated demand functions used to determine commodity prices), our dual approach also neatly integrates the restrictions on demand implicit in the use of a price index.

III. The One-Good Monetary Model

Suppose that there is one good G with a nominal price π. The investor with income I and utility function $u(G, B)$ over that good and real balances B solves the problem

$$\max_{G,B} u(G,B) \quad \text{subject to } \pi G + \pi B = I.$$

This yields the ordinary demand functions $G(\pi, I)$ and $B(\pi, I)$. The *compensated* demand functions for G and B give the demands for goods and real balances as functions of prices and the consumer's utility level u. In the one-good case, π is the "price" of both G and B so their *relative* price is always 1. Hence, at a given level of utility, a change in π has no effect on demand and the compensated demand functions for G and B can be written as functions $g(u)$ and $b(u)$ of u alone. Throughout, goods and real balances are assumed to be normal in consumption so that $g_u > 0$ and $b_u > 0$.

In state θ, π and u are determined by the conditions for equilibrium in the commodity and money markets:

$$g(u) = \theta z, \tag{7}$$

$$\pi b(u) = \mu. \tag{8}$$

Differentiating this system with respect to θ,

$$g_u u_\theta = z, \tag{9}$$

$$b\pi_\theta + \pi b_u u_\theta = 0. \tag{10}$$

By (7), (9), and (10),

$$d \log \pi / d \log \theta = \theta \pi_\theta / \pi = -gb_u / bg_u. \tag{11}$$

An immediate implication of (11) is that the real payoff on the equity is negatively correlated with inflation, a conclusion consistent with the empirical findings of Bodie (1976) and Eugene Fama (1981). Moreover, if we interpret θ as an index of real activity then Fama's explanation for this phenomenon applies in our model: real activity is correlated positively with equity returns and negatively with inflation.[7]

When $u = V(\pi, I)$,

$$B_I = b_u V_I \quad \text{and} \quad G_I = g_u V_I, \tag{12}$$

and by (11) and (12),

$$d \log(\pi\theta) / d \log \theta = 1 - \eta/\xi, \tag{13}$$

where $\eta \equiv IB_I/B$ and $\xi \equiv IG_I/G$ are the income elasticities of demand for real balances and goods. Since expenditure on these items exhausts the investor's income,

$$\eta > 1 \quad \text{if and only if } \xi < 1. \tag{14}$$

(13) and (14) yield

PROPOSITION 1: *In the one-good model with disturbances to output,*

$$1 > (<) \eta \quad \textit{implies } d(\pi\theta)/d\theta > (<) 0.$$

The intuition is as follows. An increase in θ increases the supply of the good, so π must fall to increase the demand for the good. Both the good and real balances have the same price π so the fall in π has no substitution effect and equilibration occurs solely via its income effect. If $1 > (<) \eta$, then increases in income have a relatively small (large) effect on the demand for B and therefore a relatively large (small) effect on the demand for G. Consequently, the required fall in π is relatively small (large) and the overall effect of θ on the equity payoff $\pi(\theta)\theta$ is positive (negative).

---

[7] A similar result was obtained by Theodore Day in a multi-period model under more restrictive assumptions on consumer preferences than those utilized here. In his framework, the quantity theory of money with constant velocity holds so that a 1 percent rise in output leads to a 1 percent fall in the price level. As can be seen from (11), this conclusion is obtained in our framework if and only if we impose homothetic preferences over G and B. Danthine and Donaldson also obtain a similar result in their model, which admits a nonconstant velocity of circulation. The latter authors recognized (Result 5.4, p. 565) that the correlation between the nominal payoff on the equity and real output disturbances depends on the elasticity of the output price with respect to output but they did not relate the outcome to underlying structural parameters, as we shall do in Proposition 1.

We now calculate $dV_I/d\theta$ using a method which is also valid in the n-good model. The budget constraint on consumption choices implies that

$$\pi g(V(\pi, I)) + \pi b(V(\pi, I)) = I.$$

Differentiating this with respect to I yields

$$V_I = 1/\pi(g_u + b_u), \quad \text{when } u = V(\pi, I). \tag{15}$$

Therefore,

$$V_{II}/V_I = -(g_{uu} + b_{uu})V_I/(g_u + b_u) \tag{16}$$

and

$$\begin{aligned} &d\log V_I/d\log\theta \\ &= d\log\pi/d\log\theta \\ &\quad -(g_{uu} + b_{uu})\theta u_\theta/(g_u + b_u) \end{aligned} \tag{17}$$

$$= \theta u_\theta b_u/b + \theta u_\theta V_{II}/(V_I)^2 \quad \text{by (10) and (16)} \tag{18}$$

$$= (\eta - \rho)\theta u_\theta/IV_I \quad \text{by (12)}, \tag{19}$$

where $\rho \equiv -IV_{II}/V_I$ is the investor's coefficient of relative risk aversion. Since (9) implies that $u_\theta > 0$, we have

PROPOSITION 2: *In the one-good model,*

$$\eta > (<)\rho \quad \textit{implies } dV_I/d\theta > (<)0.$$

As shown in (15), a rise in θ affects V_I in two ways: through utility u and through the price level π. The rise in θ raises u (see (9)) so, for fixed π, this reduces the marginal utility of income V_I in proportion to the investor's coefficient of relative risk aversion ρ (the second term in (17)). For fixed u, the fall in π induced by the rise in θ will raise the marginal utility of income V_I (the first term in (17)). Given a fixed nominal money supply, the extent of the fall in the price level—and hence of the rise in V_I—is proportional to the income elasticity of demand for real balances η. Proposition 2 states that the overall effect of θ is to raise the marginal utility of income if the income elasticity of demand for real balances exceeds the coefficient of relative risk aversion.

Combining Propositions 1 and 2 via Lemma 1 and invoking (6), we have

PROPOSITION 3: *In the one-good model with output disturbances,*

(A) *If* $\rho > (<)\eta > (<)1$, *then* $e < i$.

(B) *If* $1 > (<)\eta$ *and* $\rho > (<)\eta$, *then* $e > i$.

(A) confirms that there are values of the structural parameters of the economy under which the expected nominal return required on the equity will be *less* than the nominal interest rate. This unexpected outcome occurs if either nominal equity payoffs or the marginal utility of nominal income respond to real output θ in an unconventional fashion. The first case arises when $\eta > 1$ so that a rise in θ *decreases* the nominal payoff to the equity. If, in addition, $\rho > \eta$ then the dominant effect of θ on the marginal utility of nominal income V_I is via the rise in utility so V_I falls with θ. Thus, nominal equity payoffs are positively correlated with the marginal utility of nominal income and the expected nominal return required on the equity is less than that on the nominal bond—whose nominal payoffs are fixed. The second case arises when $\eta < 1$ so that a rise in θ increases the nominal payoff to the equity. If, in addition, $\rho < \eta$, then the dominant effect of θ on the marginal utility of nominal income is via the fall in the price level, so V_I *increases* with θ. Once again, there is a positive correlation between the nominal equity payoff and the marginal utility of nominal income so the expected nominal return required on the equity is less than the nominal interest rate.

This last conclusion provides an interesting perspective on equities as hedges against inflation. The conventional view is that equities serve this role because they represent claims on real assets and hence have nominal returns which are positively correlated with inflation. However, several studies have documented a negative correlation.[8] Such a correlation arises in our model when $1 > \eta$: Proposition 1 implies that θ is then positively correlated with the nominal payoff on the equity, whereas (11) implies that θ is negatively correlated with the price level. However, as Proposition 3 shows, a premium could still be paid for the equity because it provides a hedge against the shocks generating the inflation: if $\rho < \eta$ then the equity yields high payoffs when the marginal utility of income is high so that its equilibrium expected nominal return is *lower* than the nominal interest rate. Thus, by explicitly modeling the economy's response to real shocks and the investor's optimal portfolio choices in the face of these shocks, we have established that *there are a range of parameter values implying not only a negative correlation between nominal equity payoffs and inflation but also a price premium for equities as hedges against the shocks generating the inflation.*

Proposition 1 also shows that if $\eta > 1$ then the real disturbances lead to a *positive* correlation between inflation and nominal equity payoffs. However, if $\rho > \eta$, then this outcome occurs in conjunction with an equilibrium expected return on the equity which is *higher* than the nominal interest rate. Ironically, the equity is then a *poor* inflationary hedge because it yields high payoffs when the marginal utility of income is low.

In an n-good model in which the investor's preferences over goods are homothetic and the outputs of all goods are subject to perfectly correlated disturbances which leave output ratios unchanged, relative commodity prices will be unaffected by the disturbances. Hence, we effectively have a one-good, one-equity model and all the above remarks apply when we interpret θ as an index of real activity. Our analysis of the one-good model turns out to be revealing even in the general many-good case with uncorrelated disturbances in output because it captures the effects of these disturbances on the investor's choice between real balances and goods as a whole. It remains only to treat the effects arising from substitution among the various goods. This is undertaken in the following section.

IV. The Many-Good Monetary Model

Our analysis of the multi-good model exploits the restrictions on consumer preferences which are inherent in the use of a price index $\pi(p)$ to define real balances. Samuelson and S. Swamy (1974) showed that, in order for the price index to be invariant with respect to the consumer's income level, it is necessary and sufficient that his preferences over goods be homothetic. Moreover, if real balances (money balances divided by the price index) are to function as an independent argument in the utility function, then the price index must also be invariant with respect to the level of this argument, that is, consumer preferences over goods must be independent of B. Thus, we also require that the utility function U over goods and real balances be weakly separable in the goods vector X and in real balances B, that is, we require that there exist functions $G(X)$ and $u(G, B)$ such that the consumer's utility function has the form:[9]

$$U(X, B) = u(G(X), B). \tag{20}$$

Since U is homothetic in X, we can choose u and G so that G is, in fact, homogeneous of degree 1. Henceforth, we use such a representation.

The consumption problem of the investor,

$$\max_{X, B} u(G(X), B)$$

$$\text{subject to } p^k X^k + \pi(p) B = I,$$

---

[8] See the references in fn. 2.

[9] The restrictions can be relaxed by supposing that the investor's utility is a function of goods consumption, nominal money balances, and commodity prices. However, the results then become complex and difficult to interpret. Samuelson and Sato discuss this more general way of including money in the utility function and the particular functional form that we have used.

can be analyzed in two stages. First, given any expenditure I' on goods, the investor allocates I' among the various goods X^j by solving the "pure goods" problem

$$\max_X G(X) \quad \text{subject to } \Sigma p^k X^k = I'.$$

This yields the "pure goods" demand functions $X'(p, I')$, which, substituted back into G, yields the "pure goods" indirect utility function $G(X'(p, I'))$. Inverting this as a function of I', we obtain the "pure goods" expenditure function $e'(p, G)$. Since G is homogeneous of degree 1, this expenditure function has the form

$$(21) \qquad e'(p, G) = \hat{e}(p)G,$$

where $\hat{e}(p)$ is a function of p alone, as the notation indicates. (21) implies that deflating the expenditure e' on goods by $\hat{e}(p)$ yields G, the investor's real expenditure on goods. Thus, $\hat{e}(p)$ is simply the *price index* for goods—which we have previously called $\pi(p)$. The choice of a particular G and u to represent U in (20) serves to normalize this index. The upshot is that

$$(22) \qquad e'(p, G) \equiv \pi(p)G.$$

The second stage of the investor's consumption choice is to allocate expenditure optimally between goods and money. He solves

$$\max_{G,B} u(G, B)$$

$$\text{subject to} \qquad e'(p, G) + \pi(p)B = I.$$

Substituting from (22), this becomes

$$\max_{G,B} u(G, B)$$

$$\text{subject to} \qquad \pi(p)G + \pi(p)B = I.$$

Formally, this is the same problem as that treated in Section III, involving an investor with income I and utility function $u(G, B)$ over the two "goods" G and B, each with price π. The compensated demand function $g(u)$ and $b(u)$ already introduced for that problem will be used again in this new context. $b(u)$ continues to give the compensated demand function for real balances in the many-good problem.

Standard duality results[10] imply that the derivative of the expenditure function (22) with respect to p^j gives the demand for good j when the real expenditure on goods is G. Substituting $g(u)$ for G in this derivative then gives the compensated demand function for good j when his utility is u,

$$x^j(p, u) = g(u)\pi_j(p),$$

where $\pi_j \equiv \partial\pi/\partial p^j$. In state $\tilde{\theta} = (\theta^1, \ldots, \theta^n)$, the equilibrium conditions in the goods and money markets are

$$(23) \qquad g(u)\pi_j(p) = \theta^j z^j, \qquad j = 1, \ldots, n$$

$$(24) \qquad b(u)\pi(p) = \mu.$$

These $n+1$ equations determine the $n+1$ variables p^j and u. Differentiating (24) with respect to θ^j yields

$$(25) \qquad b\, d\pi/d\theta^j + \pi b_u\, du/d\theta^j = 0, \qquad j = 1, \ldots, n.$$

Thus, equations (15)–(19) in our calculation of $dV_I/d\theta$ in the one-good model remain valid for the calculation of $dV_I/d\theta^j$ in the many-good model. Moreover, it is readily checked that $du/d\theta^j > 0$. We therefore have

PROPOSITION 4: *In the many-good model with disturbances θ^j in the output of good j,*

$$\eta > (<)\rho \quad \textit{implies } dV_I/d\theta^j > (<)0, \qquad j = 1, \ldots, n.$$

We now calculate the impact of disturbances to the output of good 1 on commodity prices in the model with two goods. To simplify notation, "θ" shall be used to denote this disturbance until further notice, while $\pi_{ij} \equiv \partial^2\pi/\partial p^i\partial p^j$. Differentiating (23) and (24) with respect to θ in the two-good model yields

---

[10] See Hal Varian, 1984, p. 123.

$$(26)\quad \begin{bmatrix} g\pi_{11} & g\pi_{12} & g_u\pi_1 \\ g\pi_{21} & g\pi_{22} & g_u\pi_2 \\ b\pi_1 & b\pi_2 & b_u\pi \end{bmatrix} \begin{bmatrix} p^1_\theta \\ p^2_\theta \\ u_\theta \end{bmatrix} = \begin{bmatrix} z^1 \\ 0 \\ 0 \end{bmatrix}.$$

This system can be solved for p^1_θ, p^2_θ, and u_θ by applying Cramer's rule and simplifying the resulting determinants using the relationships among the derivatives of π arising because π is homogeneous of degree 1. To express the outcome in a tractable form, let η and ξ again denote the income elasticities of demand for real balances and goods,[11] let ε^j denote the compensated own price elasticity of demand for good j (defined so as to be positive), and let s^j be the share of goods expenditure falling on good j. To avoid confusion between superscripts and exponents, our convention is that the latter shall be used only in conjunction with brackets indicating the exponentiated term. After some simplification (see the Appendix) we have

$$(27)\quad d\log(p^1\theta)/d\log\theta$$
$$= 1 - s^1\eta/\xi - (s^2)^2/\varepsilon^1,$$

$$(28)\quad d\log p^2/d\log\theta = -s^1\eta/\xi + s^1s^2/\varepsilon^1,$$

$$(29)\quad d\log\pi/d\log\theta = -s^1\eta/\xi < 0,$$

$$(30)\quad d\log(p^1\theta/\pi)/d\log\theta = 1 - (s^2)^2/\varepsilon^1,$$

$$(31)\quad d\log(p^2/\pi)/d\log\theta = s^1s^2/\varepsilon^2 > 0,$$

$$(32)\quad du/d\log\theta = s^1IV_I/\xi > 0.$$

These equations determine the impact of the disturbance θ in the output of good 1 on the nominal and real returns to the two equities:

PROPOSITION 5: *In the two-good model*:

(A) $\{1-(s^2)^2/\varepsilon^1\}\xi/s^1 > (<)\eta$

implies $d(\theta p^1)/d\theta > (<)0$,

(B) $1-(s^2)^2/\varepsilon^1 > (<)0$

implies $d(\theta p^1/\pi)/d\theta > (<)0$,

(C) $\xi s^2/\varepsilon^1 > (<)\eta$

implies $dp^2/d\theta > (<)0$,

(D) $d(p^2/\pi)/d\theta > 0$.

The intuition for (A) is as follows. A rise in θ increases the supply of good 1. To restore equilibrium, p^1 must fall. As in the one-good case, if η is small (implying that ξ is large), then a fall in p^1 has a large income effect on the demand for goods. In addition, a change in p^1 now also has a substitution effect—which will be large if the compensated own price elasticity of demand ε^1 is large. In these circumstances, the required fall in p^1 is relatively small so an increase in θ *increases* the payoff $p^1(\theta)\theta$ on equity 1. The reverse is true if the inequalities are reversed. (B) shows that the impact of θ on the real payoff on equity 1 depends only on the strength of substitution effects, as measured by the compensated elasticity of demand ε^1: under homothetic goods preferences, income effects do not affect relative commodity prices.

Regarding (C), note that by (29), the rise in θ certainly increases the consumer's utility. If η is small (so that ξ is large) then this increase in utility leads to a strong increase in demand for good 2 via the income effect. Moreover, if ε^1 is small then the substitution effect of the fall in p^1 leads to a relatively weak decrease in the demand for good 2. Hence, the price of good 2 will tend to rise. The reverse conclusion holds if the inequalities are reversed. (D) shows that the rise in θ certainly increases the price of good 2 relative to the price level and hence the real return to equity 2: under homothetic goods preferences, income effects do not affect relative prices so the relative price of good 2 must rise in order that the increased supply of good 1 be absorbed via substitution effects.

---

[11] Our homotheticity assumption implies that all goods have the same income elasticity of demand.

Combining Propositions 4 and 5 via Lemma 1 and using (6) we have

PROPOSITION 6: *In the two-good model with disturbances in the output of good* 1,

(A) *If* $\rho > (<) \eta > (<) \{1-(s^2)^2/\varepsilon^1\} \xi/s^1$, *then the expected nominal return required on equity* 1 *is lower than the nominal interest rate.*

(B) *If* $\{1-(s^2)^2/\varepsilon^1\}\xi/s^1 > (<) \eta$ *and* $\rho > (<) \eta$, *then the expected nominal return required on equity* 1 *is higher than the nominal interest rate.*

(C) *If* $\xi s^2/\varepsilon^1 > (<) \eta > (<) \rho$, *then the expected nominal return required on equity* 2 *is lower than the nominal interest rate.*

(D) *If* $\xi s^2/\varepsilon^1 > (<) \eta$ *and* $\rho > (<) \eta$, *then the expected nominal return required on equity* 2 *is higher than the nominal interest rate.*

The rationale for Proposition 6 is a direct extension of that for the one-good model given after Proposition 3. Regarding (A), if $\eta > \{1-(s^2)^2/\varepsilon^1\}\xi/s^1$ and $\rho > \eta$, then a rise in θ reduces both the nominal payoff to equity 1 and the marginal utility of nominal income. The positive correlation between these two terms implies that the expected nominal return required on equity 1 will be lower than the nominal interest rate. The same conclusion holds when the two inequalities in the hypothesis of (A) are reversed: the rise in θ raises the nominal payoff to equity 1 but also raises the marginal utility of nominal income (via its effect on the price level). The rationales for (B)–(D) are similar.

Using the multidimensional version of Lemma 1, we can extend Proposition 6 to the case where there are uncorrelated disturbances in both outputs. The payoffs on each equity would then be a function of the disturbance vector (θ^1, θ^2). Proposition 4 gives the impact of each disturbance on the marginal utility of nominal income while Proposition 5 can be modified to give the impact of the disturbance θ^2 in the output of good 2 on the nominal payoffs on the two equities by interchanging the indices 1 and 2 for goods and equities. We refer to the resulting conclusions as Propositions 5′A, 5′B, etc. The multidimensional version of Lemma 1 can then be used to combine Propositions 4 with 5A and 5′C to obtain Proposition 6′A and 6′B below.

PROPOSITION 6′: *In the two-good model with disturbances in the outputs of both goods,*

(A) *If* $\rho > (<) \eta > (<) \{1-(s^2)^2/\varepsilon\}\xi/s^1$ *and* $\xi s^1/\varepsilon^2 < (>) \eta$, *then the expected nominal return required on equity* 1 *is lower than the nominal interest rate.*

(B) *If* $\xi s^1/\varepsilon^2 > (<) \eta$, $\{1-(s^2)^2/\varepsilon\}\xi/s^1 > (<) \eta$ *and* $\rho > (<) \eta$, *then the expected nominal return required on equity* 1 *is higher than the nominal interest rate.*

The corresponding results (C) and (D) for equity 2 can be obtained from (A) and (B) by interchanging indices 1 and 2 for goods and equities.

The above analysis permits us to extend our earlier remarks regarding equities as hedges against inflation to the case where there is more than one good and source of real disturbances.[12] Proposition 5(A) implies that if $\{1-(s^2)^2/\varepsilon^1\}\xi/s^1 > \eta$, then the disturbance θ^1 in the output of good 1 is correlated positively with the nominal payoff on equity 1 and negatively with the price level π (see (29)). Proposition 5′C implies that if $\xi s^1/\varepsilon^2 > \eta$, then the disturbance θ^2 in the output of good 2 is correlated positively with the nominal payoff on equity 1 and negatively with the price level π. The corresponding statements regarding the effects of disturbances θ^2 and θ^1 on equity 2 can be obtained by interchanging indices 1 and 2 in the above hypotheses. The resulting four hypotheses ensure a negative correlation between the nominal payoff on each equity and the price level in response to either real disturbance. If, in addition, we have $\rho < \eta$, then each equity would yield high nominal payoffs when the marginal utility of income is high and would have an equilibrium expected nominal return lower than the nominal interest rate because it provides a hedge against the shocks generating the inflation.

---

[12] If there were two goods but only one disturbance then, in the remarks that follow, the hypotheses required to determine the impact of the disturbance θ^2 would become otiose.

Similarly, by Propositions 5′B, D, and the counterparts of these propositions obtained by interchanging the indices 1 and 2, if $1 > (s^2)^2/\varepsilon^1$ and $1 > (s^1)^2/\varepsilon^2$, then the real payoff on each equity is negatively correlated with inflation in response to disturbances in either output. However, all the hypotheses invoked in the last paragraph could still hold, so both equities could still have an equilibrium expected nominal rate of return lower than the nominal interest rate. Thus, in our two-good model, Fama's argument that the observed negative correlation between real equity returns and inflation is driven by real disturbances could be valid, but this phenomenon is compatible with equities commanding a premium as hedges against the disturbances generating the inflation. *While confirming Fama's analysis, we have partially removed the puzzle which motivated that analysis.*

As in the one-good case, it is also possible to identify circumstances where real disturbances lead to a *positive* correlation between inflation and equity payoffs but where the expected nominal return on the equities is *higher* than the nominal interest rate. The extension of our analysis to n goods and equities is routine: the hypotheses determining the outcomes would merely involve $n-1$ compensated demand elasticities.

V. Monetary Disturbances

Suppose that there are no output disturbances but that, during the investment period, the investor is uncertain about what the level of the money stock μ will be during the consumption period. This might reflect uncertainty about the government's future monetary policy. Commodity prices p^j will now be a function of μ. According to (6) and Lemma 1, the expected returns on equities can be compared with the nominal interest rate by calculating the impact of μ on p^j and V_I. This is easily achieved by exploiting the homogeneity properties of the functions in the system (23) and (24) determining commodity prices. Consider a vector p and a u satisfying this system. Since $\pi(p)$ is homogeneous of degree 1 in p and $\pi_j(p)$ is homogeneous of degree 0, a 1 percent increase in μ on the right side of (24) accompanied by a 1 percent increase in each p^j would leave the system still in equilibrium. It follows that

$$(33)\qquad p^j_\mu = p^j/\mu > 0 \qquad j = 1,\dots,n.$$

The 1 percent in μ increases the consumer's income I by 1 percent also. Since $V_I(p, I)$ is homogeneous of degree -1 in (p, I), it follows that a 1 percent increase in μ *decreases* V_I by 1 percent, that is

$$(34)\qquad dV_I/d\mu = -V_I/\mu < 0.$$

Thus, an increase in the money stock increases nominal payoffs to equities just when it reduces the marginal utility of nominal income by inflation. It follows that, under monetary disturbances, the expected nominal return required on equities exceeds the nominal interest rate. We now calculate this premium explicitly.

Let $p^j(\mu)$ and $v_I(\mu)$ denote respectively the price of good j and the marginal utility of nominal income as functions of the money stock μ. Normalizing the deterministic value of each θ^j to equal 1, we have the following analogues to (5) and (6)

$$(35)\quad (i-e^j)/(1+e^j) = \frac{q^j - E[p^j(\mu)]/(1+i)}{E[p^j(\mu)]/(1+i)} = \operatorname{cov}[p^j(\mu), v_I(\mu)].$$

By (33) and (34), $p^j(\mu)$ fluctuates in proportion to μ and $v_I(\mu)$ fluctuates in proportion to $1/\mu$ so $p^j(\mu)/E[p^j(\mu)] = \mu/E[\mu]$ and $v_I(\mu)/E[v_I(\mu)] = (1/\mu)/E[1/\mu]$. Recalling that cov $[\cdot,\cdot]$ denotes the *normalized* covariance, we conclude that

$$(36)\quad \text{cov}[p^j(\mu), v_I(\mu)] = \text{cov}[\mu, 1/\mu].$$

Equations (35) and (36) imply that

$$(37)\quad (i - e^j)/(1 + e^j) = \frac{q^j - E[p^j(\mu)]/(1+i)}{E[p^j(\mu)]/(1+i)} = \text{cov}[\mu, 1/\mu].$$

Since $\text{cov}[\mu, 1/\mu] < 0$, (37) implies

PROPOSITION 7: *Under monetary disturbances, the gross expected nominal return required on each equity j exceeds that on a nominal bond by a percentage equal to the absolute value of the normalized covariance between μ and $1/\mu$. Equivalently, the price of each equity j is less than the riskless present value of its expected payoff by a percentage equal to the absolute value of this normalized covariance. This covariance is independent of consumer preferences and of j so, under monetary disturbances, the percentage rate at which the price of any equity is discounted for risk is the same for any risk-averse representative investor.*

Since (33) implies that, under monetary disturbances, the payoffs on all equities are positively correlated with inflation, it might appear that they would be good hedges against changes in inflation due to such disturbances. The above analysis shows that the reverse is true: since equities yield high nominal payoffs when the marginal utility of nominal income is low, their equilibrium expected nominal return is *higher* than the nominal interest rate.

It is a routine matter to show that under monetary disturbances, the expected real return required on an equity equals that on nominal bonds.[13]

VI. Real Bonds and the Fisher Hypothesis

Under certainty, the Fisher hypothesis states that one plus the nominal interest rate equals one plus the real interest rate multiplied by one plus the inflation rate. When inflation is uncertain, it might appear that the certainty equivalent version of this hypothesis would overstate the expected return required by risk-averse investors on real bonds, which provide automatic compensation for changes in the price level. Several authors have gone beyond this to show how the relationship between the returns on the two bonds depends on the covariance between their payoffs and the marginal utility of income. In this section, we go further still to relate this covariance to the nature of the stochastic disturbances and the structure of the underlying economy.

The Fisher hypothesis has two equivalent interpretations under certainty: that the nominal returns of real and nominal bonds are the same or that the real returns on the two bonds are the same. The question of how Fisher parity is modified by uncertainty can be approached by comparing either the nominal or the real expected returns of the two bonds. Jensen's inequality implies that the expected value of the inverse of next period's price level exceeds the inverse of the expected value of that price level, so each comparison constitutes a distinct approach to this question. Earlier authors focused on the comparison of the real returns on the two bonds.

In addition to the assets already introduced, we now suppose that there is a "real bond" whose nominal payoff increases in proportion to the price level π. If the real (rate of) return on this bond is r and the price level in the investment period is normalized to 1, then one dollar's worth of real bonds will pay $(1+r)\pi$ dollars in the consumption period. If the investor buys R dollars of real bonds, then his income in state $\tilde{\theta}$ is

---

[13] This follows immediately from Propositions 9B and 10B below.

$$I = \Sigma p^k \theta^k Z^k + N(1+i) + R(1+r)\pi + \mu,$$

and his investment problem is

$$\max_{Z^k, R, N} E\big[V\big(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k Z^k + R(1+r)\pi + N(1+i) + \mu\big)\big]$$

subject to $\Sigma q^k Z^k + N + R = W.$

The first-order condition associated with the choice of R is

$$(38)\quad E\big[\pi\big(p(\tilde{\theta})\big)V_I\big(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k Z^k + \pi R(1+r) + N(1+i) + \mu\big)\big](1+r) = \lambda.$$

Eliminating λ between (38) and (3) and assuming that the outside supply of real bonds is also zero, we conclude that the equilibrium returns on the two assets satisfy

$$(39)\quad (1+i)/(1+r) = E[\pi]\big\{1 + \mathrm{cov}\big[\pi, V_I\big(p(\tilde{\theta}), \Sigma p^k(\tilde{\theta})\theta^k z^k + \mu\big)\big]\big\}.$$

This is similar to the relationship derived by Grauer-Litzenberger (1979). We now relate the covariance in (39) to structural parameters of the economy. A routine extension of our analysis of the two-good model implies that (29) continues to hold in the n-good case, that is,

$$(40)\qquad d\pi/d\theta^j < 0.$$

Combining Proposition 4 and (39) using Lemma 1 then yields Proposition 8A below. Similarly, since (33) implies that $d\pi/d\mu > 0$, (34), (39), and Lemma 1 yield Proposition 8B.

PROPOSITION 8:

(A) *In the n-good model with uncorrelated disturbances in the output of each good j,*

$$\eta > (<) \rho \quad \textit{implies} \quad (1+r)E[\pi] > (<) + i,$$

that is, if the income elasticity of demand for real balances exceeds (is less than) the coefficient of relative risk aversion, then the gross expected nominal return required on a real bond exceeds (is less than) that on a nominal bond.

(B) *Under disturbances to the money stock μ, the gross expected nominal return required on a real bond exceeds that on a nominal bond.*

The contrasting implications of real and monetary disturbances for the direction of departures from Fisher parity highlight the importance of modeling explicitly the effects of price fluctuations on the investor's welfare and the sources of these fluctuations. Under monetary disturbances, the automatic compensation for price fluctuations provided by real bonds means that the investor receives high nominal payoffs when his marginal utility of nominal income is low. By contrast, the nominal payoffs on a nominal bond are stable. Therefore, the two bonds will be equally attractive at the margin only if the real bond yields a *higher* gross expected nominal return than a nominal bond. This can also be true if the price fluctuations are due to real disturbances. An increase in the output θ of any good reduces the price level π and hence the nominal payoff on real bonds. When $\rho < \eta$, the rise in θ will raise the marginal utility of nominal income because its effect via the reduction in the price level dominates its effect via the investor's utility. We then have a negative correlation between the marginal utility of nominal income and the nominal payoff on real bonds so their gross expected nominal return is again higher than that on nominal bonds. It is only in the case of real disturbances with $\rho > \eta$ that the gross expected nominal return on real bonds is lower than that on nominal bonds. Although this is the conclusion suggested by the simplistic argument that real bonds require a lower expected return because they provide automatic compensation for fluctuations in the price level, the above analysis shows that such automatic compensation is neither necessary nor sufficient for the conclusion: what is required is that the marginal utility of nominal income be positively correlated with fluctuations in the price level. Our analysis identifies the parameter values leading to this outcome.

We next compare the expected real returns required on the two bonds. Given our convention that the price level in the investment period is normalized to equal 1, the expected real (rate of) return $\hat{\imath}$ required on a nominal bond is given by

$$1+\hat{\imath}=E[(1+i)/\pi]. \tag{41}$$

(39) and (41) imply that[14]

$$(1+\hat{\imath})/(1+r)=E[\pi V_I]E[1/\pi]/E[V_I]. \tag{42}$$

The real returns on the two bonds can be compared by examining the correlation between their real payoffs and the marginal utility of real income πV_I under the two types of disturbances.

Let θ be the disturbance parameter in the output of any good j. The proportional impact of θ on the marginal utility of real income is

$$\begin{aligned}(43)\quad & d\log(\pi V_I)/d\log\theta \\ &= d\log\pi/d\log\theta + d\log V_I/d\log\theta \\ &= \theta u_\theta V_{II}/(V_I)^2 = -\rho s^1/\xi \\ &\qquad \text{by (16), (17), and (32).}\end{aligned}$$

Thus, if the investor is risk averse, then an increase in θ reduces the marginal utility of real income.[15] On the other hand, (40) implies that an increase in θ increases the real payoff on the nominal bond. Thus, the real payoff on a nominal bond is negatively correlated with the marginal utility of real income. Since the real payoff on a real bond is independent of the marginal utility of real income, it follows that the expected real return required on a nominal bond is higher than that on a real bond in equilibrium. Formally, by applying Lemma 1 to (40) and (43) for $\theta=\theta^1,\dots,\theta^n$ we obtain

$$E[\pi V_I]E[1/\pi]>E[V_I],$$

so by (42),

$$\hat{\imath}>r.$$

In the case of monetary disturbances, the marginal utility of real income πV_I is independent of μ so

$$\begin{aligned}(44)\quad & E[1/\pi]E[\pi V_I] \\ &\quad = E[(1/\pi)\pi V_I]=E[V_I],\end{aligned}$$

and by (42), the two assets yield the same expected real return. To summarize,

PROPOSITION 9:

(A) *Suppose that the investor is risk averse. In the n-good model with uncorrelated disturbances in the outputs of any subset of the goods, the expected real return required on a nominal bond is greater than the real interest rate.*

(B) *Under monetary disturbances, the expected real return required on a nominal bond equals the real interest rate.*

Proposition 9A can be compared to Grauer-Litzenberger's (1980) conclusion that when the monetary authority follows an automatic monetary rule in the face of exogenous shocks, the expected real return required on a real bond is less than that on a nominal bond,[16] both when the velocity of circulation of money is fixed and when it is random, but stochastically independent of the shocks. Since an automatic monetary rule renders the money supply nonstochastic, it corresponds to the situation analyzed in Proposition 9A. Our result, however, is derived from general equilibrium responses to the explicitly modeled shocks, without a priori assumptions on the velocity of circulation. Instead, this velocity is determined via the investor's maximizing response to the shocks. As in the Grauer-Litzenberger model, if the monetary authority follows a discretionary rule which stabilizes the price level in the face of output disturbances, then the payoffs on the two assets will be the same in all states so their equilibrium expected real returns will also be the same.

---

[14] Formulas similar to (42) were derived by Fama and Andre Farber, 1979; Pentti Kouri, 1983; Grauer-Litzenberger, 1980; Svensson, 1985; and LeRoy, 1984b.

[15] By contrast, Proposition 4 showed that θ has an ambiguous effect on the marginal utility of *nominal* income.

[16] These authors did not make it explicit that they were comparing the expected real returns on the two bonds but manipulation of their equation (3) shows that

It is also possible to compare the nominal and real returns required on equities with those required on real bonds. The comparison of real returns is more interesting, yielding a counter-intuitive result. Real bonds provide complete compensation for changes in the price level while, in a multi-good model, each equity can provide only partial compensation so it might appear that the expected real return required on an equity would be higher than that on a real bond. The following argument shows that this need not be the case.

Recall that by (43) the marginal utility of real income falls with θ^j when the investor is risk averse. If the real payoff on an equity also falls with θ^j, then equity j would provide a high (low) real payoff when the marginal utility of real income is high (low). By contrast, the real bond has a fixed real payoff. In these circumstances, the investor would regard the two assets as equally attractive at the margin only if the expected *real* return on the *real* bond were higher. To determine when this will occur in our two-good model with disturbances θ in the output of good 1, note that, since the price level during the investment period is normalized to equal 1, the expected real (rate of) return $\hat{e}^j$ on equity j $(j=1,2)$ is given by

$$(45)\qquad 1+\hat{e}^1 = E[p^1\theta/\pi]/q^1,$$

$$1+\hat{e}^2 = E[p^2/\pi]/q^2.$$

Therefore, (2), (38), and (45) imply that

$$(46)\qquad \frac{1+\hat{e}^1}{1+r} = \frac{E[p^1\theta/\pi]E[\pi V_I]}{E[p^1\theta V_I]}.$$

By (43), Lemma 1 and (46), if the investor is risk averse then

$$(47)\quad d(p^1\theta/\pi)/d\theta < (>)0 \quad \text{implies}$$

$$E[p^1\theta/\pi]E[\pi V_I] < (>)E[p^1\theta V_I]$$

$$\text{and}\quad \hat{e}^1 < (>)r.$$

Using Proposition 5B to relate the hypothesis of (47) to structural parameters of the economy, we obtain Proposition 10A below. For equity 2, we can use similar reasoning incorporating Proposition 5D to obtain Proposition 10B.

If there is also a disturbance θ^2 in the output of good 2, then by Proposition 5′D, a rise in θ^2 will always increase the payoff to equity 1 so we can apply the multidimensional version of Lemma 1 to obtain Proposition 10C. Proposition 10D follows from 10C by switching the indices 1 and 2.[17]

Under monetary disturbances, both the marginal utility of real income and the real payoff on any equity are fixed, so the real returns on the two assets are the same in equilibrium. This yields Proposition 10E.

PROPOSITION 10: *Suppose that the investor is risk averse. In the two-good model with*

---

this is indeed the comparison that they consider. A similar conclusion was obtained by LeRoy (1984b) and is also implicit in the work of Svensson (1985). Unlike these authors, we also compare the *nominal* expected returns on the two bonds (see Proposition 8).

[17] It is not possible to state equally general conditions implying that the expected real return required on an equity is less than the real interest rate because a higher output of one good always increases the real payoff on the other equity, even while it is decreasing the marginal utility of real income.

disturbances in the output of good 1

(A) $\varepsilon^1 <(>)(s^2)^2$ *implies* $\hat{e}^1 <(>) r$,

(B) $\hat{e}^2 > r$.

If there are disturbances in the outputs of both goods, then

(C) $\varepsilon^1 > (s^2)^2$ *implies* $\hat{e}^1 > r$,

(D) $\varepsilon^2 > (s^1)^2$ *implies* $\hat{e}^2 > r$.

Under monetary disturbances,

(E) $\hat{e}^1 = \hat{e}^2 = r$,

that is, the real returns on equities and real bonds are the same.

Notice that since a risk-averse investor's marginal utility of real income always decreases with the output disturbances θ^j, the ranking of the expected real returns on equities and real bonds depends only on the ordinal properties of the investor's consumption preferences, which determine how the real payoffs on the equities are affected by θ^j. The extension of our analysis to the case of n goods and equities is routine.

VII. Comparison of Real and Monetary Models

In Section I, we asserted that explicit modeling of money is required if we wish to compare the monetary returns on financial assets in a world where commodity prices play an economic role. It might be thought, on the contrary, that we could avoid modeling money explicitly by using a many-good model with one of the goods designated as the numeraire. However, this procedure would mis-specify the price adjustment mechanism in a monetary economy and would yield misleading results. To demonstrate this, we contrast the monetary model of this paper—in which real balances enter the consumer's utility function $U(X^1,\dots, X^n, B)$—with a nonmonetary model with the *same* utility function $U(X^1,\dots, X^n, X^{n+1})$ in which the numeraire good $n+1$ is simply an ordinary good. In such a model, the price of good $n+1$ is unity and the price of good $i<n+1$ is also its price relative to good $n+1$. By contrast, in our monetary model, the "price" of real balances is the price level $\pi(p)$ so the price of good i relative to good $n+1$ is $p^i/\pi(p)$. A fall in the price of good i will also reduce the price level and hence the "price" $\pi(p)$ of good $n+1$, so compared to the nonmonetary model, a given percentage fall in p^i will have a *smaller* effect on the price of good *i relative to that of good* $n+1$.

As a result of these differences, prices in the two systems will respond differently to output shocks. Under a disturbance to the output of good 1, p^1 must change by proportionately *more* in the monetary model in order to bring about the change in relative prices required to restore equilibrium. For this reason, less extreme conditions would be required to create a situation in which p^1 falls by enough that the payoff to equity 1 declines with an increase in θ. The stronger price movements also mean that less extreme conditions would be required to ensure that an increase in θ increases the marginal utility of income via price changes. As our discussion of Proposition 6′ showed, each of these two effects can create circumstances under which the expected nominal return required on equity 1 is lower than the nominal interest rate. Thus, a model which fails to treat money explicitly would understate the likelihood of this outcome.

This problem cannot be overcome by comparing asset returns in real terms. Such comparisons depend on the correlation between the real payoff on an asset and the marginal utility of real income. In both models, the real payoff to equity j fluctuates with the price of good j divided by the price level, that is, $p^j/\pi(p)$. In the monetary model, this is simply the price of good i in terms of the $(n+1)$th good, real balances. In the corresponding nonmonetary model, $p^j/\pi(p)$ is the price of good j in terms of good $n+1$ *divided by the price level.* Reversing our argument in the last paragraph, we see that in a nonmonetary model, the real payoffs to equity j per unit of output move more sluggishly than the price of good j relative to that of good $n+1$. Since output shocks are equilibrated by relative price movements, this means that the nonmonetary model would tend to *understate* the fluctuations in real payoffs that result from a supply disturbance. Thus, such a model would also lead to misleading conclusions about the expected real returns required on financial assets.

VIII. Extensions

In this paper, we have been able to compare the returns on various assets in terms of structural parameters of the economy within a simple framework for investor choice: consumption and investment decisions are made by a representative investor and each decision is made once. The introduction of several investors would complicate the relationship between commodity prices and output or monetary disturbances since the differential effects of prices on their portfolio incomes have to be taken into account. However, relationships between the returns required on various assets can still be inferred from the first-order conditions for optimal portfolio choice by each investor.

The extension of our model to many periods poses more challenging technical problems so we shall only foreshadow the issues. In any period of a dynamic model, the investor allocating consumption expenditure between commodities and the transaction services of real balances has an indirect utility function whose arguments include not only current consumption expenditure and commodity prices but also the opportunity cost of real balances, which now equals the interest foregone (rather than unity, as in the static model). The allocation of wealth between current consumption expenditure and investment in equities, bonds, and the store-of-value services of real balances can be expressed as a dynamic program in which the discounted expected utility achievable with a given wealth in period t is the maximized sum of the utility from current consumption expenditure and the discounted expected value of the utility achievable with the wealth left in period $t+1$, after deducting period t consumption expenditure and adding period $t+1$ dividends and the resale value of assets.

The first-order conditions for the period t maximization problem yield formulas which resemble (6): the gross nominal return on a nominal bond exceeds the gross expected nominal return required on equity j by a percentage equal to the covariance between the marginal utility of period $t+1$ wealth and the period $t+1$ return to the equity—which now comprises that period's dividends plus capital gains from resale of the equity. Lemma 1 relates this covariance to the derivatives of the equity returns and the marginal utility of wealth with respect to the random disturbances. Under rational expectations about the period $t+1$ commodity and asset prices, these derivatives can be obtained by differentiating the equations which determine these prices: the period $t+1$ equilibrium conditions for the commodity and money markets and the period $t+1$ first-order conditions for the investor's choice of equities and bonds (into which the asset equilibrium conditions have been substituted, as in the static analysis). The impact of disturbances on the marginal utility of wealth and on equity payoffs and resale values now involves feedbacks between the asset and commodity markets via the changes in the nominal interest rate and hence in the opportunity cost of real balances. The analysis of these dynamic feedbacks raises a range of issues going beyond the scope of this paper but, as we have indicated, can be pursued using similar methods.

APPENDIX

LEMMA 1: *Suppose that* $\phi(\tilde{\theta})$ *and* $\Psi(\tilde{\theta})$ *are functions of the vector* $\tilde{\theta}$ *of the independent random variables* θ^j. *If, for each* j, *either both* ϕ *and* Ψ *are increasing functions of* θ^j *or both are decreasing function of* θ^j, *then* $\text{cov}[\phi(\tilde{\theta}),\Psi(\tilde{\theta})]>0$. *If for each* j, *one of these functions is increasing and the other is decreasing in* θ^j, *then* $\text{cov}[\phi(\tilde{\theta}),\Psi(\tilde{\theta})]<0$.

PROOF:

We begin with the case where $\tilde{\theta}$ is a one-dimensional vector θ with marginal density function $f(\theta)$. Our argument for this case is a minor extension of an argument presented by James Anderson and John Riley (1976). Denote the expected values of

$\phi(\theta)$ and $\Psi(\theta)$ by $\bar{\phi}$ and $\bar{\Psi}$. The *non-normalized* covariance between $\phi(\theta)$ and $\Psi(\theta)$ is

$$\text{(A1)}\quad \int\{\phi(\theta)-\bar{\phi}\}\{\Psi(\theta)-\bar{\Psi}\}f(\theta)\,d\theta$$

$$=\int\Psi(\theta)\{\phi(\theta)-\bar{\phi}\}f(\theta)\,d\theta$$

$$=-\int\Psi_{\theta}(\theta)\int^{\theta}\{\phi(\zeta)-\bar{\theta}\}f(\zeta)\,d\zeta\,d\theta$$

upon integrating by parts.

Suppose that $\phi_{\theta}>0$. If

$$\text{(A2)}\quad \int^{\theta}\{\phi(\zeta)-\bar{\phi}\}f(\zeta)\,d\zeta\geq 0,$$

then $\phi(\theta)\geq\bar{\phi}$ and for $\zeta>\theta, \phi(\zeta)>\bar{\phi}$. It would follow that

$$0<\int_{\theta}\{\phi(\zeta)-\bar{\phi}\}f(\zeta)\,d\zeta$$

$$=-\int^{\theta}\{\phi(\zeta)-\bar{\phi}\}f(\zeta)\,d\zeta,$$

which would contradict (A2). Thus, if $\phi_{\theta}>0$ then (A2) cannot hold, that is,

$$\text{(A3)}\quad \int^{\theta}\{\phi(\zeta)-\bar{\phi}\}f(\zeta)\,d\zeta<0.$$

(A1) and (A3) imply that if $\phi_{\theta}>0$ and $\Psi_{\theta}>0$, then

$$\operatorname{cov}[\phi(\theta),\Psi(\theta)]>0.$$

The proofs of the remaining cases are similar.

Next, suppose that ϕ and Ψ are functions of the two independent random variables θ^1 and θ^2 with $\partial\phi/\partial\theta^j>0$ and $\partial\Psi/\partial\theta^j>0$, $j=1,2$. Applying the above one-dimensional result to $\phi(\theta^1,\theta^2)$ and $\Psi(\theta^1,\theta^2)$ as functions of θ^2 for a fixed value of θ^1, then taking expectations over θ^1 we obtain

$$\text{(A4)}\quad \underset{\theta^1}{E}\left[\underset{\theta^2}{E}\left[\phi(\theta^1,\theta^2)\Psi(\theta^1,\theta^2)\right]\right]$$

$$>\underset{\theta^1}{E}\left[\underset{\theta^2}{E}\left[\phi(\theta^1,\theta^2)\right]\underset{\theta^2}{E}\left[\Psi(\theta^1,\theta^2)\right]\right].$$

The independence of θ^1 and θ^2 ensures that the expectations operator over θ^2 need not be conditioned on θ^1.

Since $\partial\phi/\partial\theta^1>0$, $\partial\Psi/\partial\theta^1>0$ and θ^1 and θ^2 are independent, we have

$$\partial\underset{\theta^2}{E}\left[\phi(\theta^1,\theta^2)\right]/\partial\theta^1>0$$

and $E_{\theta^2}[\Psi(\theta^1,\theta^2)]/\partial\theta^1>0$. Applying the one-dimensional version of the Lemma to $E_{\theta^2}[\phi(\theta^1,\theta^2)]$ and $E_{\theta^2}[\Psi(\theta^1,\theta^2)]$ as functions of θ^1, we conclude that

$$\text{(A5)}\quad \underset{\theta^1}{E}\left[\underset{\theta^2}{E}\left[\phi(\theta^1,\theta^2)\right]\underset{\theta^2}{E}\left[\Psi(\theta^1,\theta^2)\right]\right]$$

$$>\underset{\theta^1}{E}\left[\underset{\theta^2}{E}\left[\phi(\theta^1,\theta^2)\right]\right]\underset{\theta^1}{E}\left[\underset{\theta^2}{E}\left[\Psi(\theta^1,\theta^2)\right]\right].$$

The conclusion then follows from (A4) and (A5) and the independence of θ^1 and θ^2. The proofs of the other cases and the extension to higher dimensions is similar.

Derivation of Equations (27)–(32). Since π is homogeneous of degree 1, Euler's theorem implies that

$$\text{(A6)}\quad p^1\pi_1+p^2\pi_2=\pi,$$

$$\text{(A7)}\quad p^1\pi_{11}+p^2\pi_{21}=0,$$

$$\text{(A8)}\quad p^1\pi_{12}+p^2\pi_{22}=0.$$

We have to apply Cramer's rule to the system (26). To evaluate the determinant Δ of the matrix on the left side, we use the fact that if $\mathbf{r}^i$ is the ith row vector of the matrix, then replacing $\mathbf{r}^1$ by $\mathbf{r}^1+\mathbf{r}^2p^2/p^1$ will not affect the determinant.

$$\Delta\equiv\begin{vmatrix} g\pi_{11} & g\pi_{12} & g_u\pi_1\\ g\pi_{21} & g\pi_{22} & g_u\pi_2\\ b\pi_1 & b\pi_2 & b_u\pi\end{vmatrix}$$

$$=\begin{vmatrix} 0 & 0 & g_u\pi/p^1\\ g\pi_{21} & g\pi_{22} & g_u\pi_2\\ b\pi_1 & b\pi_2 & b_u\pi\end{vmatrix}$$

by (A6), (A7), and (A8)

$$=(\pi_{21}\pi_2-\pi_{22}\pi_1)bgg_u\pi/p^1$$

$$=bgg_u(\pi)^2\pi_{12}/p^1p^2\qquad\text{by (A8).}$$

Let Δ_j be the determinant of the matrix obtained by substituting the right side of (26) into the jth column of the matrix on the left. By Cramer's rule

$$\text{(A9)} \quad p_\theta^1 = \Delta_1/\Delta$$

$$= \frac{z^1 p^1 p^2 \{\pi_{22}\pi g b_u - (\pi_2)^2 b g_u\}}{(\pi)^2 \pi_{12} g_u g b},$$

$$\text{(A10)} \quad p_\theta^2 = \Delta_2/\Delta$$

$$= \frac{z^1 p^1 p^2 \{\pi_1 \pi_2 b g_u - \pi\pi_{12} g b_u\}}{(\pi)^2 \pi_{12} g_u g b},$$

$$\text{(A11)} \quad u_\theta = \Delta_3/\Delta$$

$$= \frac{z^1 p^1 p^2 (\pi_{21}\pi_2 - \pi_{22}\pi_1)}{(\pi)^2 \pi_{12} g_u}$$

$$= p^1 z^1/\pi g_u \qquad \text{by (A8) and (A6).}$$

The share of goods expenditure falling on good j is

$$\text{(A12)} \qquad s^j = p^j x^j/\pi g = p^j \pi_j/\pi.$$

By (A9) and (23),

$$d \log(\theta p^1)/d \log \theta$$

$$= 1 + (\pi_1 p^2/\pi_{12}\pi^2)\{\pi_{22}\pi\eta/\xi - (\pi_2)^2\}$$

$$= 1 - s^1\eta/\xi - (s^2)^2/\varepsilon^1$$

$$\text{by (A8) and (A12)} \qquad [= (27)].$$

By (A7), (A12), and (23),

$$d \log p^2/d \log \theta = - s^1\eta/\xi + s^1 s^2/\varepsilon$$

$$[= (28)].$$

By (A9), (A10), and (A12),

$$d \log \pi/d \log \theta$$

$$= s^1 d \log p^1/d \log \theta + s^2 d \log p^2/d \log \theta$$

$$= - s^1\eta/\xi \qquad [= (29)].$$

So by (27) and (28),

$$d \log(p^1\theta/\pi)/d \log \theta$$

$$= d \log(p^1\theta)/d \log \theta - d \log \pi/d \log \theta$$

$$= 1 - (s^2)^2/\varepsilon^1 \qquad [= (30)].$$

$$d \log(p^2/\pi)/d \log \theta = s^1 s^2/\varepsilon^1 \qquad [= (31)].$$

Finally, (12), (23), (A11), and (A12) imply (32).

REFERENCES

Anderson, James E. and Riley, John G., "International Trade with Fluctuating Prices," *International Economic Review*, February 1976, *17*, 76–97.

Bodie, Zvi, "Common Stocks as a Hedge Against Inflation," *Journal of Finance*, May 1976, *31*, 459–70.

Breeden, Douglas T., "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, December 1979, *7*, 265–96.

______, "Consumption, Production, Inflation and Interest Rates: A Synthesis," *Journal of Financial Economics*, May 1986, *16*, 3–39.

Cox, John C., Ingersoll, Jonathan E., Jr. and Ross, Stephen A., "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, March 1985, *53*, 363–84.

Danthine, Jean-Pierre and Donaldson, John, "Inflation and Asset Prices in an Exchange Economy," *Econometrica*, May 1986, *54*, 585–605.

Day, Theodore E., "Real Stock Returns and Inflation," *Journal of Finance*, June 1984, *39*, 493–502.

Fama, Eugene F., "Stock Returns, Real Activity, Inflation and Money," *American Economic Review*, September 1981, *71*, 545–65.

______ **and Farber, Andre**, "Money, Bonds and Foreign Exchange," *American Economic Review*, September 1979, *69*, 639–49.

______ **and Schwert, G. William**, "Asset Returns and Inflation," *Journal of Financial*

Economics, November 1977, *5*, 115–46.

Grauer, Frederick L. A. and Litzenberger, Robert H., "The Pricing of Commodity Futures Contracts, Nominal Bonds and Other Risky Assets Under Commodity Price Uncertainty," *Journal of Finance*, March 1979, *34*, 69–83.

______ **and** ______, "Monetary Rules and the Nominal Rate of Interest Under Uncertainty," *Journal of Monetary Economics*, April 1980, *6*, 277–88.

Hansen, Lars P. and Singleton, Kenneth J., "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, September 1982, *50*, 1269–86.

Jaffe, Jeffrey and Mandelker, Gershon, "The 'Fisher Effect' for Risky Assets: An Empirical Investigation," *Journal of Finance*, May 1976, *31*, 447–58.

Kouri, Pentti J. K., "The Effect of Risk on Interest Rates: A Synthesis of the Macroeconomic and Financial Views," in *Research in International Business and Finance*, R. G. Hawkins, R. M. Levich, and C. G. Wihlborg, eds., September 1983, *3*, 301–20.

LeRoy, Stephen F., (1984a) "Nominal Prices and Interest Rates in General Equilibrium: Money Shocks," *Journal of Business*, April 1984, *57*, 77–95.

______, (1984b) "Nominal Prices and Interest Rates in General Equilibrium: Endowment Shocks," *Journal of Business*, April 1984, *57*, 197–213.

Long, John B., "Stock Prices, Inflation and the Term Structure of Interest Rates," *Journal of Financial Economics*, July 1974, *1*, 131–70.

Lucas, Robert E., Jr., "Asset Prices in an Exchange Economy," *Econometrica*, November 1978, *46*, 1429–45.

______, "Interest Rates and Currency Prices in a Two-Country World," *Journal of Monetary Economics*, November 1982, *10*, 335–59.

Merton, Robert, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, September 1973, *41*, 867–87.

Nelson, Charles R., "Inflation and Rates of Return on Common Stocks," *Journal of Finance*, May 1976, *31*, 471–83.

Rubinstein, Mark, "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, Autumn 1976, *7*, 407–25.

Samuelson, Paul A. and Sato, Ryuzo, "Unattainability of Integrability and Definiteness Conditions in the General Case of Demand for Money and Goods," *American Economic Review*, September 1984, *74*, 598–604.

______ **and Swamy, S.**, "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review*, September 1974, *64*, 567–93.

Svensson, Lars E. O., "Money and Asset Prices in a Cash-in-Advance Economy," *Journal of Political Economy*, October 1985, *93*, 919–44.

Turnovsky, Stephen J. and Brock, William A., "Time Consistency and Optimal Government Policies in a Perfect Foresight Equilibrium," *Journal of Public Economics*, April 1980, *13*, 183–212.

Varian, Hal R., *Microeconomic Analysis*, New York: W. W. Norton, 1984.

The Effects of Commodity Price Stabilization Programs

By MARIO J. MIRANDA AND PETER G. HELMBERGER*

In this paper we analyze commodity programs in which the government attempts to support and stabilize price through open market purchases and sales. Specifically, we assess the effects of such programs on the U.S. soybean market using a rational expectations model that allows for private storage, expected price-responsive production, and arbitrary support and release prices. Stochastic simulations of the model demonstrate that price stabilization programs can reduce long-run market price and destabilize producer revenue.

This paper provides an empirical evaluation of commodity price stabilization programs in which the government offers to purchase and store a commodity at a fixed support price and offers to sell any quantities in its possession at a fixed release price. Price band buffer stock programs, so-called because they are designed to contain price within the band determined by the support and release prices, have played an important role in U.S. farm policy since 1933. Such programs have been used, often in conjunction with other programs, to stabilize the domestic prices of soybeans, wheat, rice, and corn and have been promoted as a means of stabilizing international markets for a wide range of primary commodities.

Specifically, this paper assesses the effects of price band buffer stock programs on the U.S. soybean market using a rational expectations model in which acreage planted is responsive to expected future prices, yield is uncertain, and storage is carried out by competitive, expected profit-maximizing arbitrageurs. A salient feature of the model, missing from conventional models of price stabilization programs, is that private production and storage behavior adjusts to changes in government policy.[1] Rational expectations equilibria, which cannot be derived analytically, are computed numerically.

Stochastic simulations of the U.S. soybean market under alternative support and release prices produce a number of significant results regarding the effects of price band programs. We find, for instance, that only those programs entailing a low support price and a wide price band stabilize price without destabilizing producer revenue. Although this and other findings may not apply directly to other commodity or financial markets, many of our results contradict conventional wisdom and therefore have relevance beyond the narrow focus of this specific commodity. It has become commonplace in the literature, for example, to suppose that a price band program raises market price and that narrowing the price band stabilizes market price. We find that both of these suppositions are false. Some price band programs can actually reduce long-run market price below the competitive level and narrowing the band by lowering the release level can actually destabilize price.

---

*School of Business Administration, University of Connecticut, Storrs, CT 06268, and Department of Agricultural Economics, University of Wisconsin, Madison, WI 53705, respectively. We are grateful to Stephen M. Robinson, Joseph W. Glauber, and an anonymous referee for helpful comments and suggestions.

[1] The theoretical models of price stabilization following in the tradition of Frederick Waugh, 1944, Walter Oi, 1961, and Benton Massell, 1969, ignore private storage completely. The econometric studies by Jere Behrman, 1978, Olarn Chaipravat, 1978, Meghnad Desai, 1966, Shlomo Reutlinger, 1976, and Willard Cochrane and Yigal Danin, 1976, assume private stockholding and planting decisions are not affected by changes in government stabilization policy; as such, these studies are subject to Robert Lucas's 1976 critique. Regarding previous econometric research on stabilization schemes, Stephen Salant (1983, p. 2) writes, "Although many of the recent agreements and proposals have been evaluated favorably, the methodology underlying these evaluations renders their favorable conclusions worthless."

Section I sets forth the market model and discusses the methods used to compute rational expectations equilibria. Section II discusses the market effects of price band programs. Section III examines the welfare effects and government costs of price band programs. Section IV summarizes our findings.

I. The Market Model

Our annual model of the U.S. soybean market includes consumers, producers, arbitrageur storers, and the government.

Initial Conditions. A typical year t begins with a given supply s_t' held by private market participants and a given amount y_t' held by the government. Initial private supply s_t' is composed of private carryover from the preceding year x_{t-1} and new production, which equals the acreage planted the preceding year a_{t-1} times a random per acre yield $\tilde{\omega}_t$:

$$s_t' = x_{t-1} + a_{t-1} \cdot \tilde{\omega}_t. \tag{1}$$

Initial government stocks y_t' are composed of government carryover from the preceding year y_{t-1}:

$$y_t' = y_{t-1}. \tag{2}$$

Price Band Programs. The government offers to buy unlimited quantities of the commodity at a set support price $\bar{p}_s$ and to sell any quantities in its possession at a set release price $\bar{p}_R$. Government open market operations alter the distribution of total supply between private and government hands; denoting *net* government purchases in year t by g_t, final private supply is given by

$$s_t = s_t' - g_t, \tag{3}$$

and final government stocks by

$$y_t = y_t' + g_t. \tag{4}$$

The government does not purchase stocks if the market price exceeds the support level and does not sell any if the market price falls below the release level

$$p_t > \bar{p}_s \Rightarrow g_t \leq 0, \tag{5}$$

$$p_t < \bar{p}_R \Rightarrow g_t \geq 0. \tag{6}$$

Since the government is willing to acquire any amount of the commodity at the support price, market price never falls below the support level

$$p_t \geq \bar{p}_s. \tag{7}$$

The government, however, is able to release only as much as it holds in its stockpile initially,

$$g_t \geq -y_t'. \tag{8}$$

Thus, market price can rise above the release level, though only if the government stockpile is depleted:

$$p_t > \bar{p}_R \Rightarrow g_t = -y_t'. \tag{9}$$

Private Market Activity. Of final private supply, an amount x_t is stored by arbitrageurs and the remainder, $s_t - x_t$, is purchased by consumers at the market clearing price

$$p_t = \pi(s_t - x_t). \tag{10}$$

Competition among the expected profit-maximizing arbitrageurs eliminates profit opportunities; this yields the well-known intertemporal arbitrage conditions[2]

$$\begin{aligned} &x_t \geq 0, \\ &p_t \geq \delta p_{t+1}^e - \kappa, \\ &p_t > \delta p_{t+1}^e - \kappa \Rightarrow x_t = 0. \end{aligned} \tag{11}$$

---

[2] Given the widespread use of forward and futures contracts and the growing evidence that risk premiums may be small, we judge the present assumption of risk neutrality provides a useful first approximation. See Jeffrey Frankel (1984) for further discussion of this point.

Here, $\delta p^e_{t+1} - \kappa$, the discounted price expected next year less the constant unit storage cost, is the expected marginal revenue from storing the commodity.

The acreage planted by producers depends on the price expected next year at harvest time,[3]

$$a_t = \alpha(p^e_{t+1}). \tag{12}$$

Arbitrageurs and producers form their expectations of next year's price rationally,

$$p^e_{t+1} = \mathscr{E}_t p_{t+1}. \tag{13}$$

Model Parameterization. Farm-level demand and acreage supply equations for U.S. soybeans were estimated using ordinary least squares and 1951–80 data.[4] Exogenous shifters were assigned their 1977 values to obtain simplified exponential forms expressing quantity demanded in billions of bushels, and acreage supplied, in millions of acres, as functions of market price and expected market price, respectively, in real (1977) dollars per bushel. The simplified demand equation had a constant term of 5.18 and an elasticity of -0.61; the simplified acreage supply equation had a constant term of 13.0 and an elasticity of 0.89.

The random yields $\tilde{\omega}_t$ were assumed to be independently and identically distributed following a log-normal distribution with an estimated mean of 29.84 bushels per acre and a coefficient of variation of 17.43 percent. The annual discount factor of 0.916 was derived from a real rate of interest estimated by taking the interest rate on loans for feeder cattle adjusted for the change in the producer price index. Historical data suggested an approximate storage cost of $0.36 per bushel per year.

Rational Expectations Equilibria. To solve the market model (1)–(13), one must determine how expected price varies with changes in other endogenous variables. Stephen Salant (1983) and Miranda (1987) have shown that if price expectations are rational and the model is time stationary, then there exists an unique function f such that

$$\mathscr{E}_t p_{t+1} = f(x_t, y_t, a_t). \tag{14}$$

That is, the rationally expected price is uniquely determined by the current levels of private carryout, government carryout, and acreage planted. Given the conditional expected price function f, substituting (14) for (13) closes the model, allowing it to be solved for the equilibrium levels of price, private carryout, government carryout, and other endogenous variables.

The function f cannot generally be derived analytically.[5] However, an approximation to the function f can be numerically computed through successive approximation: Begin with an arbitrary guess f_0 for f. Having the nth iterate f_n, substitute

$$p^e_{t+1} = f_n(x_t, y_t, a_t) \tag{15}$$

for (13) and solve equations (3)–(13) for the equilibrium price for each level of initial private supply and government stocks on some specified grid. Interpolate the grid values bilinearly to construct a function ϕ_n that expresses the approximate equilibrium price

---

[3] The market model can easily be generalized to include stochastic terms in the demand and acreage supply equations; see Mario Miranda (1985). Preliminary simulation experiments indicated that the effects of these random shocks are negligible in comparison to the effects of yield variability. They were dropped to reduce the cost of computation.

[4] The demand equation was estimated using semiannual data with the demand elasticity restricted to be the same for both halves of the year. The semiannual demands were summed to obtain annual demand. In estimating the acreage supply function, the November futures price for soybeans as of planting time, deflated by an index of prices paid for farm inputs, was used as a proxy for expected price. The t-ratio for the demand elasticity was 2.19; the t-ratio for the supply elasticity was 3.65.

[5] In many recent econometric rational expectations macroeconomic studies (for example, Ray Fair and John Taylor, 1983), the conditional expected price relationship is embedded in the econometric model and is estimated directly. In our study, the conditional expected price function does not appear in the structural econometric model. The econometric model is estimated independently first and only then is the conditional expected price function implied by the model computed numerically.

$\phi_n(s', y')$ for all possible levels of initial private supply s' and government stocks y'. Construct the $n+1$th iterate by setting

$$(16)\quad f_{n+1}(x_t, y_t, a_t) = \mathscr{E}\tilde{\omega}\phi_n(x_t + y_t + \tilde{\omega}\cdot a_t, y_t)$$

for all x_t, y_t, and a_t. If the difference between f_n and f_{n+1} is sufficiently small, stop; otherwise repeat the steps above.

The algorithm above typically converged in six iterations or less to an approximation for the rational expected price function exhibiting a relative error not exceeding one-tenth of one percent. For details see Miranda (1987).[6]

II. Market Effects of Price Band Programs

Analysis of the effects of price band programs on the U.S. soybean market is based on the comparison of the steady-state means and coefficients of variation of selected market variables under different price band regimes. For each regime, the steady-state moments were estimated from samples of two million observations generated through Monte Carlo simulations of the market process. The simulations were performed by repeatedly drawing random yields from the assumed log-normal distribution.

Table 1 provides estimates of the steady-state mean government stocks, mean price, coefficient of variation of price, and coefficient of variation of producer revenue for various combinations of support and release price. Explosive policies, those whose steady-state mean government stocks are infinite, are marked with asterisks. Although not evident from this table, policies whose support price exceeds \$5.65 are explosive, regardless of the width of the band. Lowering the support price widens the range of price bands consistent with finite steady-state mean government stocks. Roughly speaking, the range for nonexplosive bands grows by forty cents for every twenty-cent drop in the support price.

Our first major result is that nonexplosive price band policies lower steady-state mean market price below its competitive level of \$5.72 (see Table 1). For this class of policies, steady-state mean price falls monotonically if either the support level or the bandwidth is increased. Only explosive price band policies raise the steady-state mean price above the competitive level. For explosive policies, steady-state mean price rises monotonically if either the support price or the bandwidth is increased.

A second major result evident from Table 1 is that price can be stabilized substantially without causing government stocks to explode. For example, setting the support and release prices at \$5.50 and \$5.70, respectively, lowers the steady-state coefficient of variation of price from 20.24 percent, the competitive level, to 5.15 percent. The associated steady-state mean government stock equals 1.88 billion bushels, roughly one year's production. In general, raising the support price while holding bandwidth constant stabilizes market price. Perfect price stabilization can be achieved in the long run only by selecting an explosive policy with a zero bandwidth.

Price band programs affect price behavior by altering market demand and supply. Figure 1A gives the total demand for consumption and private storage, DD_C, under competitive conditions. Introduction of a price band program immediately causes total demand to become perfectly elastic at the support level $\bar{p}_s$, the result of the government's willingness to acquire unlimited stocks at that price. Total demand becomes DD_B. In the first year of the program, total supply, which is predetermined and thus perfectly inelastic, is composed of private carryover from the preceding year and new production. For a low market supply s_0, price p_0 exceeds the support level, obviating the need for government open market purchases. For a high market supply s_1, on the other hand,

---

[6] Brian Wright and Jeffrey Williams, 1982, and Mark Lowry et al., 1985, use computable rational expectations models to study competitive storage. Bruce Gardner, 1979, and Gerald Plato and Douglas Gordon, 1983, use similar methods to study government buffer stock schemes. These studies draw on the work of Robert Gustafson, 1958, who was the first to use computational methods to study commodity storage.

TABLE 1—STEADY-STATE MEAN PRICE, COEFFICIENT OF VARIATION OF PRICE, COEFFICIENT OF VARIATION OF PRODUCER REVENUE, AND MEAN GOVERNMENT STOCKS[a, b]

Support Price		Bandwidth (Release Price Minus Support Price)						
		0.00	0.20	0.40	0.60	0.80	1.00	1.20
4.50	Mean Price	5.71	5.71	5.71	5.71	5.71	5.71	5.71
	CV Price	20.01	19.98	19.89	19.74	19.58	19.45	19.38
	CV Revenue	8.70	8.69	8.63	8.47	8.28	8.08	7.92
	Govt. Stocks	0.03	0.03	0.03	0.04	0.04	0.04	0.04
4.70	Mean Price	5.71	5.71	5.71	5.70	5.70	5.70	5.70
	CV Price	19.44	19.28	18.95	18.54	18.09	17.58	17.01
	CV Revenue	9.51	9.43	9.16	8.82	8.45	8.05	7.62
	Govt. Stocks	0.06	0.06	0.07	0.08	0.10	0.13	0.17
4.90	Mean Price	5.70	5.70	5.69	5.69	5.69	5.68	5.68
	CV Price	18.19	17.69	17.05	16.29	15.42	14.37	13.23
	CV Revenue	10.68	10.34	9.89	9.40	8.87	8.30	7.67
	Govt. Stocks	0.09	0.11	0.13	0.16	0.21	0.30	0.53
5.10	Mean Price	5.69	5.68	5.68	5.67	5.67	5.66	5.70
	CV Price	16.04	15.14	14.05	12.75	11.16	9.21	9.39
	CV Revenue	11.85	11.33	10.75	10.14	9.48	8.77	7.95
	Govt. Stocks	0.14	0.17	0.22	0.32	0.58	2.43	***
5.30	Mean Price	5.67	5.67	5.66	5.66	5.70	5.75	5.80
	CV Price	13.03	11.49	9.53	6.83	6.52	7.89	9.12
	CV Revenue	13.07	12.42	11.74	11.01	10.06	9.20	8.53
	Govt. Stocks	0.24	0.33	0.58	2.26	***	***	***
5.50	Mean Price	5.66	5.65	5.70	5.77	5.82	5.86	5.90
	CV Price	8.62	5.15	3.39	4.90	6.27	7.49	8.57
	CV Revenue	14.32	13.57	12.52	11.44	10.58	9.91	9.41
	Govt. Stocks	0.55	1.88	***	***	***	***	***
5.70	Mean Price	5.70	5.79	5.86	5.91	5.95	5.98	6.01
	CV Price	0.00	1.68	3.21	4.58	5.79	6.86	7.79
	CV Revenue	15.24	13.93	12.85	11.99	11.32	10.82	10.45
	Govt. Stocks	***	***	***	***	***	***	***
5.90	Mean Price	5.90	5.97	6.02	6.06	6.09	6.11	6.13
	CV Price	0.00	1.55	2.91	4.11	5.15	6.05	6.83
	CV Revenue	15.24	14.15	13.30	12.64	12.15	11.79	11.53
	Govt. Stocks	***	***	***	***	***	***	***

[a] Prices are expressed in real (1977) dollars per bushel, coefficients of variation are expressed as percentages, and government stocks are expressed in billions of bushels. Asterisks denote infinite steady-state government stocks.

[b] For a competitive regime in steady state, mean price equals $5.72 per bushel, coefficient of variation of price equals 20.24 percent, and coefficient of variation of producer revenue equals 8.30 percent.

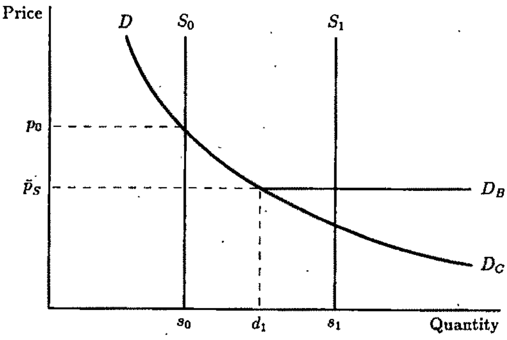


FIGURE 1A. DEMAND AND SUPPLY UNDER A PRICE BAND PROGRAM $t=1$

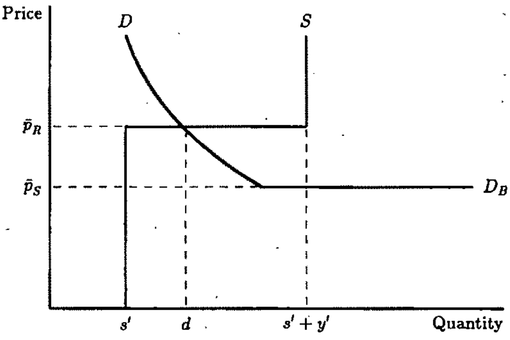


FIGURE 1B. DEMAND AND SUPPLY UNDER A PRICE BAND PROGRAM $t>1$

the government must withdraw an amount $s_1 - d_1$ from the open market to prevent competitive forces from driving price below the support level.

After the first year, the government will typically have accumulated a stockpile through earlier open market purchases. As illustrated in Figure 1B, the government's offer to sell its stocks at the release price $\bar{p}_R$ alters open market supply. At prices below the release level, open market supply S equals s', the sum of private carryover from the preceding year and new production. At prices equal to or exceeding the release price, open market supply is displaced rightward by the amount offered by the government, say, y'. For a low initial private supply such as s', the government sells an amount $d' - s'$ on the open market to prevent competitive market forces from driving price above the release level.

The overall effect of a price band program on price level and price variability can thus be decomposed into two distinct effects. The demand effect, arising from the government's stock acquisition policy, is realized immediately upon the introduction of the program and acts to raise market price. The supply effect, arising from the government's stock release policy, varies over time with the level of government stocks and acts to reduce market price. As seen in Table 1, for certain price band programs, the supply effect eventually dominates the demand effect, causing steady-state mean market price to fall below the competitive level. The demand and supply effects work together to stabilize price. The former prevents sharp price downswings, the latter prevents or otherwise moderates sharp upswings.

Figure 2 illustrates the dynamic, short-term impact of introducing a price band program into the U.S. soybean market. Specifically, this figure depicts market performance during the fifteen years following the introduction of a \$5.10–\$5.50 price band program into a previously competitive market.[7] In the first year of the program, the government on average acquires substantial stocks, raising mean market price and reducing price variability by preventing price downswings. In subsequent years, the mean level of government stocks approaches its steady-state value (Figure 2A). The growth of government stocks puts downward pressure on price, moderating the price hikes realized in the first year of the program (Figure 2B). By the fifth year, mean market price has dropped below the competitive steady-state price of \$5.72. The growth of government stocks enhances the government's ability to prevent sharp price upswings, further stabilizing price. As seen in Figure 2C, substantial price stability is realized soon after the introduction of a price band program.

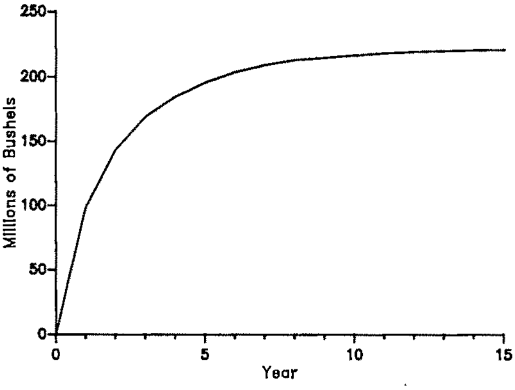


FIGURE 2A. MEAN GOVERNMENT STOCKS UNDER A \$5.10–\$5.50 PRICE BAND PROGRAM

A third major result from Table 1 is that for nonexplosive policies, widening the price band by raising the release price stabilizes market price, contradicting the claims of several writers (Walter Labys, 1978; Willard Cochrane and Yigal Danin, 1976; Olarn Chaipravat, 1978; and James Langley et al.,

---

[7] The means and coefficient of variation presented in Figure 2 were estimated from a sample of forty thousand simulated market histories. It is assumed that producers and arbitrageurs do not anticipate the introduction of the program. Miranda (1985) performs a similar analysis using the same model but under the alternative scenario that the government announces the program a year in advance. Comparing the results, we conclude that the effects of the announcement are negligible.

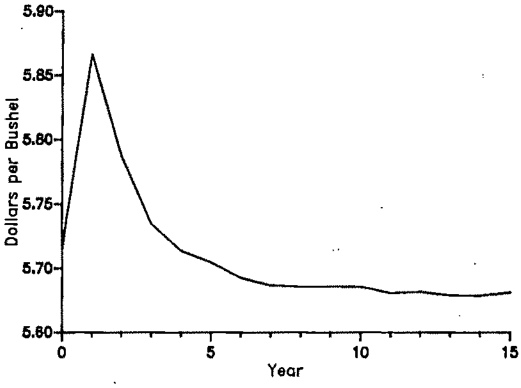


FIGURE 2B. MEAN MARKET PRICE UNDER A $5.10–$5.50 PRICE BAND PROGRAM

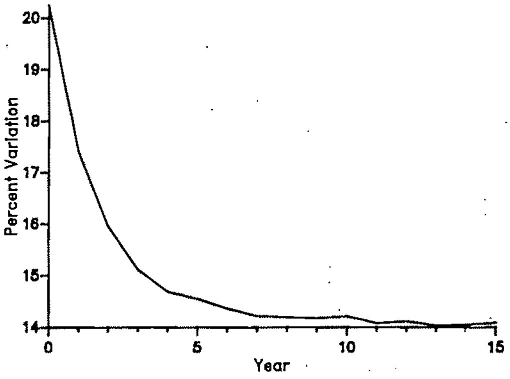


FIGURE 2C. COEFFICIENT OF VARIATION OF MARKET PRICE UNDER A $5.10–$5.50 PRICE BAND PROGRAM

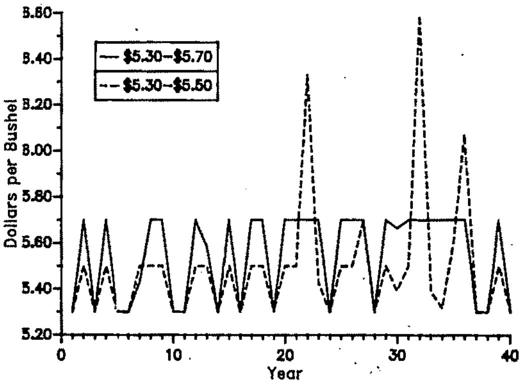


FIGURE 3. SIMULATED HISTORIES OF MARKET PRICE UNDER $5.30–$5.50 AND $5.30–$5.70 PRICE BAND PROGRAMS

1985). Raising the release price leads to a larger long-run level of government stocks, reducing the likelihood of high prices resulting from government stock-outs during strings of poor harvest. In other words, it is easier to contain price within a wide price band than to contain it within a narrow one. For explosive policies, on the other hand, government stock-outs cease to occur in the long run, causing the release price to become an effective price ceiling; widening the band destabilizes price.

Figure 3 illustrates how bandwidth affects price variability in the case of nonexplosive policies. This figure shows representative price histories for price band programs with a $5.30 support price and bandwidths of $0.20 and $0.40, respectively; both histories were generated by the same sequence of exogenous supply shocks. Over the first twenty years, price is contained within the price bands and exhibits greater variability under the wider band. The remainder of the forty-year period, by contrast, is characterized by a sequence of poor harvests. For the narrow band, the government buffer stock is quickly depleted, resulting in a series of high prices. For the wide band, on the other hand, the government has a larger stockpile when the series of poor harvests begins, enabling it to contain price within the band throughout the period of poor harvests. We estimate that in the long-run government stock-outs occur with probability 0.24 for the narrow band regime and probability 0.14 for the wide band regime.

A fourth result is of interest in light of the frequent assumption that support and release prices can be maintained equidistantly from a trend price. For example, Jere Behrman (1978, p. 307) assumes that "the buffer stock managers operate with sufficient financial and commodity reserves so that they can buy or sell to keep the average annual deflated commodity prices within specified bands (±15 percent) of the known secular trends...." Such strategies may be infeasible. For many nonexplosive policies, as shown in Table 1, the entire price band lies below the long-run price. Bracketing a given target price with arbitrary equidistant support and release prices does not guarantee that prices will subsequently fluctuate symmetrically about the target.

A fifth result evident from Table 1 is that a simple relationship between price and revenue variability does not exist. For nonexplosive price band policies, raising the support price stabilizes price but destabilizes revenue. For the same class of policies, however, raising the release price stabilizes both price and revenue. Importantly, only price band policies involving low price supports and wide bands reduce the coefficient of variation of total revenue below its competitive level of 8.3 percent. Even in these cases, the reductions are small. Most price band policies destabilize total revenue.[8]

Price band programs affect revenue variability by altering the relation between price and production. Under competitive conditions, if production falls, price rises, and vice versa; price response offsets variations in production, tending to stabilize producer revenue. Price band programs destabilize producer revenue by suppressing price response to production variation. For example, increasing the support price from $4.50 to $5.50, holding bandwidth equal to zero, causes the steady-state correlation between price and production to rise from −0.83 to −0.36. Price becomes less responsive to supply conditions and producer revenue is destabilized even though price is stabilized. Holding the support price at $5.10 but increasing the bandwidth from $0.00 to $1.00 causes the correlation between price and production to fall from −0.63 to −0.84. Widening the band enhances price responsiveness, helping to stabilize revenue.

The variation in total revenue is important because of its implications for the variation of producer quasi rent. We do not report coefficients of variation for producer quasi rent because their measurement involves using the entire area under the supply curve to estimate cost; this requires placing reliance on a portion of the estimated supply curve well outside the range of observed data. Nonetheless, it is probably significant that the variability of producer quasi rent follows the same pattern as that of revenue. For example, introducing a price band program with a $5.30 support price and a $0.20 band into a competitive market causes the coefficient of variation of total revenue to rise from 8.3 percent to 12.4 percent. For this same change in regime, the coefficient of variation of quasi rent rises from 13.2 percent to 22.5 percent. Summing up, price band programs cannot be relied upon to stabilize either producer revenue or quasi rent. On the contrary, such programs often have the opposite effect.

To test the sensitivity of our findings to demand and supply elasticity, we performed simulations for which the elasticities were varied 20 percent above and below their base values (namely, 0.89 for supply and −0.61 for demand). Table 2 gives selected results of these simulations for three different regimes: a competitive market, a price band program for which the support and release prices both equal $5.30, and a price band program for which the support and release prices equal $5.30 and $5.50, respectively.

From Table 2 we see that the patterns found in Table 1 are preserved as elasticities vary. For all elasticities, the selected price band programs lower steady-state mean price, reduce long-run price variability, and increase long-run revenue variability. Moreover, comparing the results for the $5.30–$5.30 price band to those of the $5.30–$5.50 price band, we see that raising the release price reduces and stabilizes market price and destabilizes revenue, regardless of the elasticity.

Table 2 also shows how the impact of a price band program varies with the elasticities of demand and supply. First, the decline in steady-state mean price resulting from the introduction of a nonexplosive price band program, though small, is more pronounced as either demand or supply becomes more inelastic. Second, price and revenue variability appear to be insensitive to supply elasticity, but highly sensitive to demand elasticity.

---

[8] Several writers (for example, Jere Behrman, 1978, and David Newbery and Joseph Stiglitz, 1981) have examined the relationship between price and revenue variability, though without regard to the underlying stabilization mechanism.

TABLE 2—STEADY-STATE MEAN PRICE, COEFFICIENT OF VARIATION OF PRICE, AND COEFFICIENT OF VARIATION OF PRODUCER REVENUE UNDER COMPETITION AND TWO ALTERNATIVE PRICE BAND PROGRAMS FOR VARIOUS COMBINATIONS OF DEMAND AND SUPPLY ELASTICITIES[a]

		Competitive Market			\$5.30–\$5.30 Price Band			\$5.30–\$5.50 Price Band		
	Demand Elasticity	Supply Elasticity 0.71	0.89	1.06	0.71	0.89	1.06	0.71	0.89	1.06
Mean Price	−0.49	5.73	5.72	5.71	5.68	5.67	5.67	5.67	5.67	5.66
	−0.61	5.73	5.72	5.71	5.68	5.67	5.67	5.67	5.67	5.66
	−0.74	5.72	5.71	5.70	5.68	5.67	5.67	5.67	5.67	5.67
CV Price	−0.49	23.9	23.6	23.3	15.0	14.7	14.5	13.1	12.9	12.6
	−0.61	20.4	20.2	20.1	13.2	13.0	12.8	11.7	11.5	11.3
	−0.74	17.9	17.7	17.6	11.9	11.8	11.6	10.6	10.5	10.3
CV Revenue	−0.49	11.6	11.6	11.6	14.2	14.2	14.3	13.3	13.3	13.4
	−0.61	8.2	8.3	8.4	12.9	13.0	13.2	12.3	12.4	12.5
	−0.74	5.8	6.0	6.1	12.1	12.2	12.3	11.7	11.8	11.9

[a] In the base case, demand and supply elasticities equaled −0.61 and 0.89, respectively.

Third, the destabilizing effects of price band programs on revenue are more pronounced as demand becomes more elastic. Fourth, both price and revenue variability are more sensitive to demand elasticity under competition than under price band programs, reflecting the decline in the relative importance of private consumption demand under a price band program.

A number of other results, which we now summarize, are not apparent from either Table 1 or 2. Nonexplosive policies, which reduce long-run price, also reduce long-run production and consumption. Although such policies scarcely affect production variability, they can stabilize consumption substantially. For example, a \$5.10–\$5.50 price band policy lowers the steady-state coefficient of variation of consumption from its competitive value of 11.1 percent to 7.2 percent while leaving the coefficient of variation of production unchanged at 15.6 percent. We also find that even price band policies that have only a modest effect on price variability can have a profound impact on private stockholding. For example, if the support price is set equal to \$4.90 and the bandwidth is set equal to 80 cents, steady-state mean private stocks plummet from a competitive level of 65 million bushels to about two million bushels.[9]

III. Some Benefit-Cost Considerations

The dynamic welfare effects of introducing a price band program to a historically competitive market are illustrated in Figure 4. This figure gives the expected changes in consumer and producer surplus and the expected government outlays in each of the first fifteen years of a \$5.10–\$5.50 price band program.[10] As seen in Figure 4, producer gains and consumer losses are substantial in the first year of the program but subsequently moderate. This is consistent with Figure 2A, which shows that introducing a \$5.10–\$5.50 price band program causes mean price to rise sharply at first and then to fall gradually toward its steady-state value. As seen in Figure 4, government outlays in the first year of the program are high due to the

---

[9] Jimmye Hillman, D. Gale Johnson, and Roger Gray (1975, p. 14) urge that the "range between acquisition and selling prices must be large enough to permit private firms to obtain profits from holding stocks." They provide no reason for believing that widening the price band would encourage private stockholding, however. Our analysis suggests that for a given price support, widening the band has practically no effect on private stockholding. The scope of private stockholding depends almost exclusively on the support level.

[10] Producer surplus in year t is defined as $CS_t = p_t \cdot h_t - \delta \cdot c(a_{t-1})$, where h_t is quantity harvested in year t and the cost of planting in year $t-1$ is given by

$$c(a_{t-1}) = \mathscr{E}_{t-1}\tilde{\omega}_t \cdot \left[\mathscr{E}_{t-1} p_t \cdot a_{t-1} - \int_0^{\mathscr{E}_{t-1} p_t} \alpha(p^e)\, dp^e \right].$$

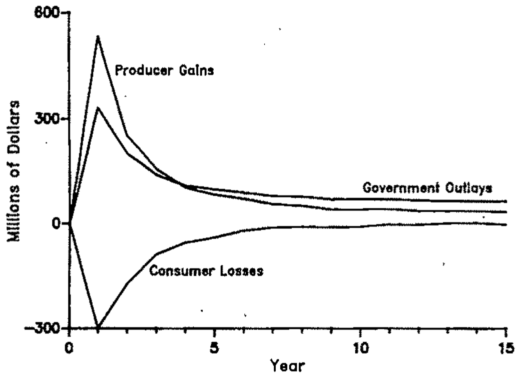


FIGURE 4. DYNAMIC WELFARE EFFECTS OF A $5.10–$5.50 PRICE BAND PROGRAM

large government stock acquisitions and the absence of government sales. Government outlays subsequently fall but remain positive, even though mean government net open market purchases approach zero, due to the costs of storing the buffer stocks.

Most comparative static studies focus on the steady-state welfare effects of price stabilization and on the sensitivity of these effects to demand and supply curvature. Introducing a price band program into a competitive market invariably leads to immediate and substantial producer gains and consumers' losses that greatly exceed the smaller and more heavily discounted long-run effects. Because comparative static studies fail to recognize the importance of the short-run welfare effects, their results are potentially misleading.[11] Moreover, since the elevation of expected price is inevitable in the short run, quite without regard to the curvature of demand or supply, the importance these studies attach to curvature may be exaggerated.

Table 3 gives for various price band programs the average annual producer gains and consumer losses expected over the fifty years following the introduction of the program into a previously competitive market.[12] As is evident in Table 3, producers gain and consumers lose from the introduction of price band programs. That this is true even for policies that reduce long-run market price is perhaps surprising. As indicated earlier, however, the effects of low prices realized in the distant future are heavily discounted and have a small impact on the present value of the sums of producer and consumer surplus. For the most part, welfare gains and losses are determined in the early years of the program when prices are high. Consumer losses during the first five years of a $5.10–$5.50 price band program, for example, account for 89 percent of the total losses incurred over the first fifty years; producer gains during the first five years account for 63 percent of the gains obtained over the first fifty years.

Table 3 also gives expected annual government outlays and expected annual deadweight loss (consumer losses plus government outlays less producer gains) over the first fifty years of program operation. These estimates indicate that substantial price stability can be obtained for relatively modest government outlays. A $5.10–$5.50 price band program, for example. reduces price variability by about one-third soon after it is introduced (see Figure 1B). The cost of such a program equals, on average, $115 million annually, a meager sum compared to the billions of dollars spent annually on conventional farm programs.

Figure 5 illustrates how one might achieve both price and revenue stability at minimal cost. In this figure, the steady-state coefficients of variation of price and revenue are plotted against the discounted sum of gov-

---

[11] In their analysis of competitive storage, Wright and Williams (1984) also recognize the dynamic welfare effects of the introduction of storage. The impact from the introduction of government storage, however, can be significantly greater than the impact from the introduction of competitive storage. Mean steady-state stocks under price band program with a $5.50 support level and a zero bandwidth, for example, is ten times what it is under competitive conditions.

[12] Producer gains and consumer losses were measured by the change in producer and consumer surplus and were discounted to produce an amortized average annual figure for each of the ten thousand market histories simulated. It was assumed that the introduction of the program was not anticipated by producers or arbitrageurs.

TABLE 3—AVERAGE ANNUAL PRODUCER GAINS, CONSUMER LOSSES, GOVERNMENT OUTLAYS, AND DEADWEIGHT LOSSES, IN MILLIONS OF DOLLARS, EXPECTED FROM INTRODUCTION OF A PRICE BAND PROGRAM TO A COMPETITIVE MARKET[a]

Support Price		Bandwidth (Release Price Minus Support Price)						
		0.00	0.20	0.40	0.60	0.80	1.00	1.20
4.50	Producer Gains	7	7	9	10	12	12	12
	Consumer Losses	4	5	6	6	9	9	10
	Govt. Outlays	5	6	6	7	9	10	10
	Deadweight Loss	3	3	3	4	6	7	8
4.70	Producer Gains	22	24	27	31	33	39	46
	Consumer Losses	9	11	13	16	18	23	32
	Govt. Outlays	17	18	21	26	33	44	59
	Deadweight Loss	4	5	7	11	18	29	45
4.90	Producer Gains	52	56	59	66	76	90	105
	Consumer Losses	21	24	26	32	43	59	77
	Govt. Outlays	40	44	53	67	88	118	163
	Deadweight Loss	10	13	20	33	55	87	135
5.10	Producer Gains	94	102	113	128	148	173	201
	Consumer Losses	39	46	57	74	97	127	162
	Govt. Outlays	80	93	115	152	207	281	367
	Deadweight Loss	25	37	59	97	156	235	328
5.30	Producer Gains	153	171	195	227	265	304	339
	Consumer Losses	70	88	116	155	201	250	294
	Govt. Outlays	151	190	254	348	462	581	692
	Deadweight Loss	68	107	175	275	398	527	647
5.50	Producer Gains	246	287	337	390	438	478	511
	Consumer Losses	133	181	243	307	367	417	457
	Govt. Outlays	305	419	569	728	877	1004	1109
	Deadweight Loss	193	314	474	646	805	942	1055
5.70	Producer Gains	417	488	553	607	650	683	710
	Consumer Losses	285	371	450	516	568	610	643
	Govt. Outlays	687	900	1097	1263	1399	1508	1595
	Deadweight Loss	555	783	994	1172	1318	1434	1527
5.90	Producer Gains	687	757	813	857	890	917	937
	Consumer Losses	546	631	698	751	792	823	848
	Govt. Outlays	1360	1577	1750	1887	1995	2079	2145
	Deadweight Loss	1220	1451	1635	1781	1896	1985	2055

[a] Amortized averages over the first fifty years of program operation. Producer gains and consumer losses are measured by the changes in producer and consumer surplus, respectively.

ernment outlays for policies involving bandwidths of $0.00 and $1.20. The figure shows that for any government outlay policies of either bandwidth provide essentially the same degree of price stability. The figure also shows that for any government outlay the wider band provides substantially more revenue stability. These results, which hold for other bandwidths as well, suggest that if price and revenue stability are both important goals, then an efficient policy will entail a low price support and a wide band.

Two caveats regarding the recommendation of a low price support and a wide price band are in order. First, the degree of revenue stability that can be secured through price band programs is at best limited, as we have already noted. If the primary objective of government intervention is revenue stabilization, then the effectiveness of price band programs is questionable. Second, certain levels of price stability may be unattainable without destabilizing revenue or causing government stocks to explode. As shown in Table 1, any nonexplosive price band policy that reduces the coefficient of variation of price below 10.1 percent, half its competitive value, also destabilizes total revenue.

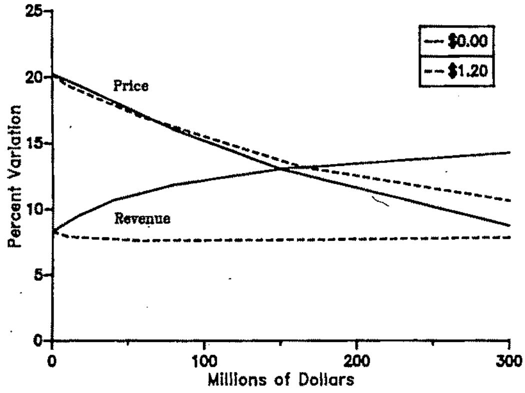


FIGURE 5. PRICE AND REVENUE VARIABILITY VS. EXPECTED ANNUAL GOVERNMENT OUTLAYS FOR PROGRAMS WITH BANDWIDTHS OF $0.00 AND $1.20

IV. Summary

In this paper we have examined a rational expectations model of the U.S. soybean market in which the government attempts to stabilize price within a specified band through open market purchases and sales. The model extends conventional static equilibrium models to a dynamic framework, allowing for private storage, expected price-responsive production, exhaustible government stocks, and distinct support and release prices.

Through stochastic simulations of the market model we obtained a number of novel results regarding the effects of price band buffer stock programs. First, although all price band programs raise market price in the short run, some can also reduce market price in the long run. Only price band policies that lead to explosive government stocks effect a sustainable increase in market price. Second, although price band programs can substantially stabilize price, they can also destabilize producer revenue. Only policies with low support prices and wide bands stabilize both price and producer revenue, and even then the degree of revenue stability attainable is relatively modest. Third, widening the price band by raising the release price lowers price variability in the case of nonexplosive policies. Narrow price bands are more difficult to defend because they are more susceptible to stock-outs. And finally, we have shown that the welfare effects of price band programs are determined largely in the first few years of the program with producers gaining and consumers and taxpayers losing.

Numerical solution and simulation of a rational expectations model has been shown in this paper to be a powerful technique for analyzing intertemporal equilibrium models too complex to handle analytically. Although computable rational expectations models are difficult and expensive to develop, the effort and cost involved in implementing them may well be worthwhile, particularly where essential market features might otherwise be ignored to ensure mathematical tractability.

REFERENCES

Behrman, Jere R., "International Commodity Agreements: An Evaluation of the UNCTAD Integrated Programme for Commodities," in F. Gerard Adams and Sonia A. Klein, eds., *Stabilizing World Commodity Markets*, Lexington: Lexington Books, 1978.

Chaipravat, Olarn, "International Rice Buffer Stock Operations: A Simulation Study," in F. Gerard Adams and Sonia A. Klein, eds., *Stabilizing World Commodity Markets*, Lexington: Lexington Books, 1978.

Cochrane, Willard W. and Danin, Yigal, *Reserve Stock Grain Models, the World and the United States, 1975–1985,* Minnesota Agricultural Experiment Station Technical Bulletin, No. 305, University of Minnesota, Saint Paul, 1976.

Desai, Meghnad, "An Econometric Model of the World Tin Economy, 1948–1961," *Econometrica*, January 1966, *34*, 5–134.

Fair, Ray C. and Taylor, John B., "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models," *Econometrica*, July 1983, *51*, 1169–85.

Frankel, Jeffrey A., "Commodity Prices and Money: Lessons from International Trade," *American Journal of Agricultural*

Economics, December 1984, *66*, 560–66.

Gardner, Bruce L., *Optimal Stockpiling of Grain*, Lexington: Lexington Books, 1979.

Gustafson, Robert L., *Carryover Levels for Grains: A Method for Determining Amounts That Are Optimal Under Specified Conditions*, Technical Bulletin No. 1178, United States Department of Agriculture, Washington: USGPO, October 1958.

Hillman, Jimmye, Johnson, D. Gale and Gray, Roger, *Food Reserve Policies for World Food Security: A Consultant Study on Alternative Approaches*, Food and Agricultural Organization of the United Nations, Rome, ESC:CSP/75/2, January 1975.

Labys, Walter C., "A Critique of Econometric Studies of Commodity Price Stabilization," *Proceedings of the Business and Statistics Section*, American Statistical Association, August 1978, 187–93.

Langley, James A., Reinsel, Robert D., Craven, John A., Zellner, James A. and Nelson, Frederick J., "Commodity Price and Income Support Policies in Perspective," in U.S. Department of Agriculture, Economic Research Service, *Agricultural-Food Policy Review: Commodity Program Perspectives*, Agricultural Economics Report No. 530, U.S. Department of Agriculture, Economic Research Service, July 1985.

Lowry, Mark N., Glauber, Joseph W., Miranda, Mario J. and Helmberger, Peter G., "Pricing and Storage of Field Crops: A Quarterly Model Applied to Soybeans," *American Journal of Agricultural Economics*, November 1987, *69*, 740–49.

Lucas, Robert E., Jr., "Econometric Policy Evaluation: A Critique," in K. Brunner and A. Meltzer, eds., Carnegie-Rochester Conference on Public Policy, *Journal of Monetary Economics*, Suppl. 1976, *1*, 19–46.

Massell, Benton F., "Price Stabilization and Welfare," *Quarterly Journal of Economics*, May 1969, *83*, 284–98.

Miranda, Mario J., "A Computable Rational Expectations Model for Agricultural Price Stabilization Programs," unpublished paper, University of Connecticut, Storrs, 1987.

______, "Analysis of Rational Expectations Models for Storable Commodities Under Government Regulation," unpublished doctoral dissertation, University of Wisconsin-Madison, 1985.

Newbery, David M. G. and Stiglitz, Joseph E., *The Theory of Commodity Price Stabilization*, Oxford: Oxford University Press, 1981.

Oi, Walter Y., "The Desirability of Price Instability Under Perfect Competition," *Econometrica*, January 1961, *29*, 58–64.

Plato, Gerald and Gordon, Douglas, "Dynamic Programming and the Economics of Optimal Grain Storage," *Agricultural Economics Research*, January 1983, *35*, 10–22.

Reutlinger, Shlomo, "A Simulating Model for Evaluation Worldwide Buffer Stocks of Wheat," *American Journal of Agricultural Economics*, February 1976, *58*, 1–12.

Salant, Stephen W., "The Vulnerability of Price Stabilization Schemes to Speculative Attack," *Journal of Political Economy*, February 1983, *91*, 1–38.

Waugh, Frederick V., "Does the Consumer Benefit from Price Instability?," *Quarterly Journal of Economics*, August 1944, *58*, 602–14.

Wright, Brian D., and Jeffrey C. Williams, "The Economic Role of Commodity Storage," *Economic Journal*, September 1982, *92*, 596–616.

______ **and** ______, "The Welfare Effects of the Introduction of Storage," *Quarterly Journal of Economics*, February 1984, *99*, 169–92.

Counterfeit-Product Trade

By GENE M. GROSSMAN AND CARL SHAPIRO*

We analyze trade in both legitimate and counterfeit products. Domestic firms own trademarks and establish reputations for delivering high-quality products. Foreign suppliers export legitimate low-quality merchandise and counterfeits of domestic brand-name goods. Home consumers have rational expectations regarding counterfeiting. We describe the positive and normative effects of counterfeiting, and provide a welfare analysis of border inspection policy and of policy regarding the disposition of counterfeit goods that are confiscated at the border.

Trade in counterfeit products is reaching epidemic proportions. Casual observers are becoming increasingly aware of the presence of fakes and trademark-infringing knockoffs in the markets for a wide variety of products, including not only the traditionally forged, luxury consumer goods such as designer clothing, watches, perfumes, and leather items, but also higher-technology consumer electronic products such as computers and stereo equipment. There is also mounting evidence of substantial counterfeiting in the markets for records and tapes, foods, pharmaceuticals, and an expanding range of industrial goods, including parts for automobiles and airplanes, fertilizers, pesticides, military hardware, and medical devices.[1] *Business Week*, in a recent cover story devoted to counterfeiting (December 16, 1985), called it "perhaps the world's fastest growing and most profitable business."

The growth in counterfeit-product trade has attracted the attention of the international trade community, including corporations, governments, and the international organizations. Firms are hiring specialized private detective services to track down, expose, and prosecute the forgers (Jack Kaikati and Raymond LaGarce, 1980), and are devoting more and more resources to making their brand-name products copy-proof (Sandra Salmans, 1979). A number of companies adversely affected by counterfeiting have bound together to form such organizations as the International Anticounterfeiting Coalition in New York and the Union des Fabricates in Paris. These associations lobby governments for stricter domestic laws, tighter border control, and tougher sanctions against countries that foster illegitimate producers. The governments, for their part, are beginning to heed the call. For example, the U.S. Congress passed the Trademark Counterfeiting Act of 1984 and the Tariff and Trade Act of 1984 after extensive congressional hearings on the topic, and the U.S. International Trade Commission (USITC) conducted a comprehensive investigation of counterfeiting in 1983 with a view toward reforming its procedures for handling complaints under the relevant section of the Tariff Act of 1930. Most recently, the United States, Japan, and the European community have agreed to make the adoption of an anti-counterfeiting code one of their primary objectives for the proposed

---

*Department of Economics and Woodrow Wilson School, Princeton University, Princeton, NJ 08544. We are grateful to Sandy Grossman and Michael Katz for their help in refining the equilibrium concept; and to Franklin Allen, Avinash Dixit, Jean Baldwin Grossman, and seminar participants at places too numerous to mention for useful comments on an earlier draft. We also thank the National Science Foundation under grant no. SES-8606336, the Alfred P. Sloan Foundation, and the Office of Foreign Economic Research, U.S. Department of Labor for financial support.

[1] Anecdotal evidence on the extent of counterfeiting and the range of products affected appears regularly in the public and business presses. See the many articles cited in Shari Olenick (1982) and the U.S. House of Representatives Committee on Energy and Commerce (1983, 1984). Data from a survey of U.S. businesses are available in the USITC (1984).

upcoming round of GATT negotiations (*New York Times*, January 20, 1986).

The U.S. International Trade Commission (1984, p. vii) defines counterfeiting as "the unauthorized use of a registered trademark on a product that is identical or similar to the product for which the trademark is registered and used." The anti-counterfeiting code drafted by GATT goes further in ascribing to the forger the intent to "wrongfully benefit through *deceit* from the efforts of a firm to establish and maintain a product or corporate image with the consumer or the public at large" (emphasis added). Counterfeiting, like patent and copyright infringement, represents a violation of a firm's property rights, in this case the rights to its trademark and associated goodwill. It is distinguished from these related practices, however, in that it alone involves an attempt to defraud consumers via misrepresentation.

Deceptive counterfeiting arises in markets with imperfectly informed consumers. If consumers could immediately and costlessly observe all of the attributes of goods available for purchases, it would be impossible for imitators to pass off products of inferior quality under false labels. Counterfeiters might nonetheless make positive sales in such situations if they offered their goods at discounted prices, and if consumers valued the exclusivity and prestige attached to conspicuous consumption of certain well-known labels. We study nondeceptive counterfeiting of status goods in our companion paper (forthcoming). Here we focus on the effects of counterfeiting in markets where consumers cannot detect bogus goods, either by direct inspection or by inference from the place of purchase.[2]

Trademarks play an important role in a world of imperfect information. When a firm invests in its reputation by delivering a promised quality, it develops goodwill with its customers. Trademarks allow consumers to identify the products of companies that have satisfied them in the past. Thus, a trademark becomes an asset of the firm, embodying its accumulated goodwill. When governments grant firms exclusive property rights to their marks, they protect firms' investments. Without such protection, firms would find it difficult to appropriate the benefits from maintaining the quality of their products and would have less incentive to do so.

Counterfeiting undermines the functioning of the property rights system. Not only do consumers suffer the direct harm associated with the purchase of low-quality copies purporting to be originals, but the infringement on the legitimate firms' rights alters the incentives to invest in their reputations. It does so in two ways. First, consumers will be willing to pay less for high-quality products in situations in which they recognize a risk of obtaining fakes. Second, a consumer who purchases a bogus good may not identify it as such, and may attribute its poor performance to the trademark holder. Then the presence of counterfeits in a market can tarnish the images of honest manufacturers.

In principle, counterfeiting need not be a trade issue. In practice, however, most counterfeits originate in certain countries in which the laws governing the protection of trademarks are not so strict and enforcement is lax. Indeed, in many less-developed countries, the importation of foreign technologies is a conscious development strategy, and the line between imitation and infringement sometimes becomes blurred. U.S. producers surveyed by the USITC identified Taiwan as the source country in 60 percent of the cases in which they experienced competition from counterfeit products (USITC, 1984). Other countries implicated as havens for firms producing counterfeits include Hong Kong, Indonesia, South Korea, the Philippines, and Thailand. Italy supplies many of the counterfeits sold in Europe, while a number of Middle Eastern countries are fast becoming prominent in this market.

---

[2] In practice, it may not be possible to characterize markets for counterfeit products as being either wholly deceptive or wholly nondeceptive. Consumers may differ in their abilities to distinguish fakes, and distribution outlets may vary in the reputations they have developed for carrying authentic products. Nonetheless, we feel that the analytical distinction is a useful one for highlighting the different implications of these two sorts of counterfeit-product trade.

In this paper, we develop an equilibrium model of counterfeit-product trade. In doing so, we incorporate into our analysis not only the direct effects of counterfeiting that arise when a consumer purchases a fake instead of a genuine brand-name product, but also the induced effects on the behavior of legitimate producers. We pay special attention to the quality-choice decision of brand-name firms, recognizing that consumers will only purchase from companies that make credible offers and deliver on their promises.

In Section I we develop our model, which takes as its starting point the burgeoning literature on equilibrium in markets with imperfectly informed consumers.[3] To set a benchmark for comparison, we begin Section II by establishing the properties of the equilibrium under the assumption that counterfeiting is not feasible. We then introduce the possibility that some imports may be fakes, ask when counterfeiting will occur, and use comparative static techniques to study the determinants of the market share of counterfeits. We conclude this section by comparing home and world welfare levels to those in the benchmark case. In Section III we study the welfare effects of border inspection policy and of customs policy regarding the disposition of confiscated products. Finally, a concluding section contains a summary of the results, some of which are quite surprising.

I. A Model of Counterfeit-Product Trade

We consider the world market for "blue jeans," a product of heterogeneous quality. The attributes of blue jeans that determine their quality are not immediately observable by consumers. Two types of producers are active in the blue jeans market. Home-country firms have well-developed, quality-control procedures, and the ability and incentive to invest in their reputations. These are the "name brand" producers. Foreign producers lack the capability to produce high-quality merchandise; but these firms have comparative advantage in the production of low-quality (henceforth, "generic") blue jeans. The location of production by the home-country firms (i.e., whether or not direct foreign investment takes place) is irrelevant to our analysis.

Our model is cast in discrete time with an infinite number of periods of length T. Each home firm has its own distinguishable trademark, but otherwise jeans of like quality produced by different manufacturers are perfect substitutes. Variable production costs for home firms are $xc(q)$, where x is output and $c(q)$ is the constant marginal cost of producing jeans of quality q, with $c' > 0$ and $c'' > 0$. Implicitly, we are assuming that these firms have access to perfectly elastic supplies of not only labor, but also such other factors as managers and capital, which they may purchase on the (large) home market. Each home firm also faces an "entry cost" F, which it must incur one period before engaging in any production. This fee can represent, for example, the cost to the firm of developing its product or advertising its trademark.

Per-period profits of a representative home firm are $\pi = (p - c(q))x - rF$, where $r \equiv e^{iT} - 1$ (i is the instantaneous discount rate) and thus rF is the interest cost on the entry fee.[4] This expression presumes, as will be the case in equilibrium, that the firm sells all that it produces. We entertain two alternative assumptions about market structure: the number of home firms M is fixed, or there is free entry. In the event of the latter, M adjusts at the beginning of time until there are no excess profits. In all cases, we assume that M is large.

The foreign industry produces only blue jeans of some minimal quality q_0, because foreign entrepreneurs are assumed to lack access to the technology for quality control. In keeping with the stylized facts, we suppose that foreign firms are small, that entry is free to foreign entrepreneurs, and that the technology exhibits constant returns to scale.[5] But we assume that resources such as capital and entrepreneurial talent are in scarce supply in the foreign country, so that as foreign production of blue jeans expands, the prices of these factors are driven higher. Denoting aggregate production by foreign entrepreneurs by X^*, foreign firms' unit production costs are given by $c^*(X^*)$ with $c^{*\prime} > 0$. Each small foreign firm treats c^* as a constant.

---

[3] See, for example, Benjamin Klein and Keith Leffler, 1981; Carl Shapiro 1982, 1983; and Franklin Allen, 1984. In our model, we follow Shapiro (1983) in assuming that firms can choose from a range of possible qualities, while we borrow from Allen the assumption that consumers form their expectations rationally.

[4] These are profits in a systematic, steady-state equilibrium. Under the assumptions we make about expectation formation and consumer shopping strategies, no asymmetric equilibria exist and adjustment to the steady state is instantaneous.

At the beginning of each period, each home firm selects a price p, a quality q, and announces (for example, advertises) the price and a quality "claim" $\hat{q}$. The true quality must meet or exceed q_0, because we assume that consumers can identify items that will fail to "do the job." After the announcements have been made, home and foreign firms select their output levels simultaneously, and offer their goods for sale in the home-country market.[6]

In marketing their output, foreign firms must choose how to label their products. If a pair of foreign jeans is labeled honestly, consumers infer that it is of quality q_0. Let p_0 denote the price of such goods. A foreign firm can instead mimic the label of a home producer. A pair so labeled is a *counterfeit*. If the foreign firm is successful in selling such a pair, it receives the price p attached to high-quality jeans. However, first the firm must pass the goods through the home-country border. There, it faces the risk of confiscation by home-country customs agents.[7] We assume that confiscation occurs with probability ϕ, where $\phi \equiv bf(\hat{q})$.

ϕ is the product of two components. The first component b is the probability that a given package is opened. We consider b to be a home-country policy parameter, reflecting the intensity of border inspection. The second component is the probability of detection conditional on a given shipment being inspected. In the United States and elsewhere, trademark owners supply customs agents with information not readily available to consumers about the likely recognizable differences between counterfeit and authentic goods, and with samples of the latter. The agents use the information and the samples provided in their search for counterfeits. We assume that the greater the difference between the quality of the sample and that of the counterfeit, the more likely that the agents will identify the copies as such. Suppose that the samples are made to match the announced quality $\hat{q}$, and that customs agents accept the samples as legitimate if the announcement is credible. Then we can write the conditional probability of detection, given q_0, as $f(\hat{q})$, with $f' > 0$.

Turning to the demand side of the model, we assume that all consumers purchase one unit of blue jeans. We distinguish two types of consumers. Quality-conscious consumers, who number N in total, value a unit of quality at θ in terms of the numeraire. That is, the total utility each enjoys by consuming one unit of jeans of quality q purchased at price p is $U=\theta q-p+y$ for $q \geq q_0$, where y is income and $y-p$ is expenditure on the numeraire good. The remaining N_0 home consumers value quality less highly, perhaps because their incomes are smaller. These consumers are happy to purchase a low-quality product, as long as it will do the job. The form of their utility functions is the same as for quality-conscious consumers, but with y and θ replaced by y_0 and θ_0, where $\theta_0<\theta$. We take parameter values so that, at the prices prevailing in equilibrium, this group of consumers most prefers the generic product among those that are (or could be) offered.

---

[5] According to an attorney quoted in the *Business Week* article, in the counterfeiting business "there's no Mr. Big. It's a bunch of little guys." The article later explains that "it is now incredibly cheap to make fakes that once required factories with heavy equipment and hundreds of workers. Today one man in an auto repair shop can copy the contours of a fender on a home computer. He can then make a plastic die and run off hundreds of copies."

[6] We assume, without consequence for any of our results, that no sales take place in the foreign market.

[7] The commercial laws of many countries call for the confiscation of trademark-infringing goods (see USITC, 1984). In the United States, two statutes apply. Under the provisions of the Lanham Act, a trademark owner may record his *registered* mark with the Customs Service, which then will prohibit entry of goods bearing counterfeit marks. The Tariff Act of 1930 calls for seizure and forfeiture of infringing goods. A firm claiming infringement of its *common law* trademark by imports may apply to the International Trade Commission (ITC) for relief under Section 337 of the Tariff Act of 1930. If its claim is validated and a determination of injury is made, the Commission will issue an exclusion order to the Customs Service. Customs then has the choice whether or not to permit reexport of goods seized under such an order.

As in Shapiro (1983), consumers cannot immediately observe the characteristics of a product that set it above the minimum quality, q_0. Furthermore, warranties for jeans are assumed to be unenforceable. Hence, a quality-conscious consumer who chooses other than the generic product buys a good of uncertain worth. If the person expects that counterfeits account for a fraction s of the goods carrying a particular brand-name label and that the legitimate product is of quality $\hat{q}$, then expected utility from purchasing this brand is $U=s(\theta q_0-p+y)+(1-s)(\theta\hat{q}-p+y)=\theta[sq_0+(1-s)\hat{q}]-p+y$. The consumer does learn the quality of a product upon consuming it, that is, with a lag of one period. However, if the product turns out to be of quality q_0, the consumer never learns whether the good was, in fact, a counterfeit.

We assume that consumers also observe the total quantity of jeans of each label available on the market.[8] If consumers could not observe quantity, a legitimate firm that promised a better deal than its rivals could, if believed, flood the market with minimum-quality goods. At high levels of sales, the quick profits associated with being "fly-by-night" exceed the sustainable profits of an honest firm that maintains its reputation. Consumers would be skeptical to an extreme of any "good deals," and would buy only from firms promising the prevailing levels of surplus, whatever that level happened to be. If, however, consumers can observe total sales, or at least be aware when the market is being flooded by a particular brand, then a firm can credibly offer a good deal by restricting its output, thereby preserving its incentive to deliver on its promise.

Finally, we specify how consumers form their expectations and choose their brands. Each quality-conscious consumer accepts at face value the quality claim of any firm, so long as the claim is *credible* and the consumer has not been disappointed by a product bearing the particular label in the past. The consumer assesses credibility by determining whether, if believed, the firm has incentive to deliver on its promise. In equilibrium, all firms produce their advertised quality. Thus, consumers' expectations about quality are "rational."

Consumers also must form expectations about the share of counterfeits of each brand. Each consumer first calculations whether, at the prevailing price and (expected) quality, entry by counterfeiters is deterred, or whether counterfeits should be expected to flood the market. Then, if neither of these extremes is believed to be the case, the consumer assumes that the market share of counterfeits in the relevant submarket will be "normal," that is, it will be such that the brand in question yields the same expected utility as do others. In a symmetric equilibrium, these expectations, too, are fulfilled.

Consumers calculate the expected surplus associated with each brand, using their subjective probabilities. Each initially chooses the brand offering the highest surplus. If several brands offer the same expected surplus, selection among these is random. After purchasing, the consumer observes whether the quality of the item is equal to that claimed by the trademark owner. If so, and if no other brand offers higher expected surplus in the next period, the consumer buys from the same company again. If, however, the jeans do not perform at quality $\hat{q}$, the consumer realizes that either the legitimate firm has shaved its quality, or else a counterfeit has been purchased inadvertently.

---

[8] In Allen's model, the equilibrium with observable quantity is also an equilibrium when quantity is not observable. The same is true in our model. However, Allen does not ask whether any new equilibria can arise once quantity becomes unobservable. In fact, there, as here, a continuum of equilibria exists when quantity is unobservable, and for much the same reason.

The consumer believes that if the legitimate firm has cheated, it will continue to do so by offering minimum quality in the future.[9] Given this belief, if the consumer could be sure that his good was authentic (albeit substandard), he would strictly prefer to switch brands. If instead the person could know that the good was in fact a counterfeit, then, at least in equilibrium, the consumer would be indifferent to switching. We can assume, therefore, without requiring any consumer irrationality, that whenever a consumer is disappointed by a purchase, he "plays it safe" and opts to try a new brand. A firm that cheats will lose all its customers and will be forced to shut down. One that lives up to its reputation will lose only those patrons who are deceived by counterfeiters of its label. In a symmetric equilibrium, this loss is just balanced period-by-period by new clients who have had unhappy experiences with their own former brands.

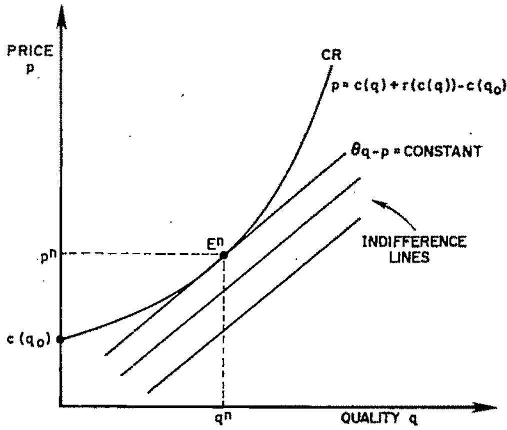


FIGURE 1. NO-COUNTERFEITING EQUILIBRIUM

II. Properties of Equilibrium and Regime Comparisons

A. *Equilibrium without Counterfeiting*

As a benchmark for comparison, we first establish the properties of the equilibrium when counterfeiting is not feasible. Formally, we assume in this subsection that $\phi \equiv 1$. Consumers infer immediately that $s = 0$ for all brands.

A home-country firm advertising price p, claiming quality $\hat{q}$, producing output x, and remaining honest earns a flow profit (gross of entry costs) of $(p - c(\hat{q}))x$. The present discounted value of a stream of such profits is $(1+r)(p - c(\hat{q}))x/r$. Alternatively, the firm could claim $\hat{q}$, but produce a lower quality. If it did so, it would be discovered after one period, and would then be forced out of business. A firm that elects to cheat maximizes its one-period profit by producing minimum quality, thereby earning $(p - c(q_0))x$. The *credibility constraint* requires that the firm have an incentive to provide its announced quality. This gives

$$p \geq c(q) + r(c(q) - c(q_0)). \tag{1}$$

As Shapiro (1983) has argued, promises of above-minimum quality can be credible only if firms earn premia over their marginal costs. These premia provide the incentives for firms to maintain their reputations.

Given $p \geq c(q)$, firms surely will compete for customers. Competition forces the credibility constraint to equality, or

$$p^n = c(q^n) + r(c(q^n) - c(q_0)), \tag{2}$$

where we use a superscript n to denote equilibrium values in the no-counterfeiting regime. In Figure 1, we depict (2) as the curve *CR*.

Home producers also can try to expand sales by offering more attractive (but credible) price-quality packages than their rivals. If all firms offer the same expected surplus, then each is rationed and sells to N/M consumers. A firm that offers a slightly better deal can greatly increase its sales and therefore its profits. Since a single firm is small in the overall market, it can expand its sales by a large amount in percentage terms without giving its rivals reason to respond in the ensuing subgame. Consumers therefore can rationally expect the deviant's better deal to persist. In equilibrium, there cannot exist opportunities for profitable deviations. Thus, competition drives firms to the credible price-quality pair that maximizes quality-conscious consumers' utility.

---

[9] We will find that, if a firm decides to cheat at all, it is optimal for it to produce minimum quality. Thus, consumers who suspect that their firm might not have delivered on its promise will also find that their purchase is of quality q_0. Our assumption, then, is that consumers expect quality deviations to persist indefinitely. Among the set of equilibria that can arise under alternative assumptions about the persistence of deviations, the one we choose to study is Pareto dominant. The approach taken here also is adopted by Klein and Leffler, 1981; and Allen, 1984.

In Figure 1, we depict several indifference lines for the representative quality-conscious consumer, with utility rising as we move to the southeast. The price and quality that maximize consumers' utility subject to the credibility constraint are found at the point of tangency between the *CR* curve and an indifference line. Algebraically, we have the condition

$$c'(q^n)(1+r)=\theta, \tag{3}$$

which we term the *competition equation.*

Having determined the equilibrium price and quality of home-country jeans, the rest of the no-counterfeiting equilibrium is readily described. Each home firm is rationed in equilibrium, producing output $x^n = N/M$. Each earns per-period profits of

$$\pi^n = \left(p^n - c(q^n)\right)N/M - rF. \tag{4}$$

If entry is free, then $\pi^n = 0$, which, together with (4), determines the number of home firms, M^n.

Foreign firms serve consumers of generic jeans. Market clearing in this submarket requires

$$X^{*n} = N_0. \tag{5}$$

Free entry abroad ensures

$$p_0^n = c^*(X^{*n}). \tag{6}$$

Our assumption that foreign producers have comparative advantage in the market for low-quality jeans requires that $c^*(X^{*n}) < c(q_0)$.

In equilibrium, quality-conscious consumers receive lower-quality jeans than in the "first best," where no information problems exist.[10] But, under conditions of free entry, the no-counterfeiting equilibrium does achieve a constrained (or second-best) social optimum. If we allow the government to subsidize costs but not to administer entry taxes or provide information, then it must maximize total surplus subject to the zero-profit and credibility constraints. The first-order conditions for this maximization imply $\theta = c'(q)(1+r)$ or $q = q^n$. Given the information structure, changes in quality away from q^n are beneficial only if entry by home firms can be controlled.

B. *Equilibrium with Potential Counterfeiting*

We seek to characterize equilibrium when counterfeiting is feasible. Let us provisionally assume that copies account for a positive share of the market for each brand. Later, we will check whether this assumption is justified.

Home firms continue to face a credibility constraint in announcing their qualities. Starting from a candidate equilibrium with sales rationed, a firm that cheats on quality can sell $(1-s)N/M$ units of the minimum-quality good for one period. Alternatively, it can sell this same quantity indefinitely by fulfilling its promises. Evidently, the presence of counterfeits has no effect on the *relative* profitability of the alternative strategies. Thus, the incentive to produce the announced quality is exactly the same as in the no-counterfeiting case. As before, competition drives the credibility constraint to equality, so we have

$$p = c(q) + r\left(c(q) - c(q_0)\right). \tag{7}$$

---

[10] The first best occurs where the marginal cost of quality equals the marginal consumer valuation, or $c'(q) = \theta$. Shapiro (1982) proves and discusses a more general version of the result that imperfect consumer information leads to quality deterioration.

Firms again compete for customers, since at the margin it remains profitable to sell more jeans. But now, in their attempts to (credibly) offer surplus in excess of that offered by other brands, the legitimate producers must take into account the response of counterfeiters. A legitimate firm has three choices in this regard. If it chooses a price-quality pair such that counterfeiters can earn positive profits, consumers will expect the market for this brand to be overrun by copies. No sales result under this strategy, so it cannot be optimal. Second, a home firm can announce a price and quality at which counterfeits of its trademark just break-even. Consumers then expect a normal share of fakes, and the legitimate firm enjoys a proportionate share of total jeans sales. Finally, the firm in question can select a price and quality such that foreign producers will not wish to counterfeit its label. By doing so, the firm makes its output especially attractive to consumers, and may thereby relax its sales constraint. It captures both the share of its own submarket otherwise held by counterfeiters, and some of the market of its legitimate competitors.[11]

In any candidate equilibrium, all foreign firms (counterfeiters or otherwise) earn zero profits. Since any particular home firm is small, it takes the costs of the foreign firms as given. The home firm can drive the counterfeiters out of its submarket by changing its price-quality vector slightly in an appropriate direction. Such a deviation must be profitable, as long as it does not cause the firm to violate the credibility constraint. But profitable deviations cannot exist in equilibrium. It follows that, as a condition of equilibrium with positive counterfeiting, all price and quality changes that cause counterfeiters to earn negative profits must not be credible.

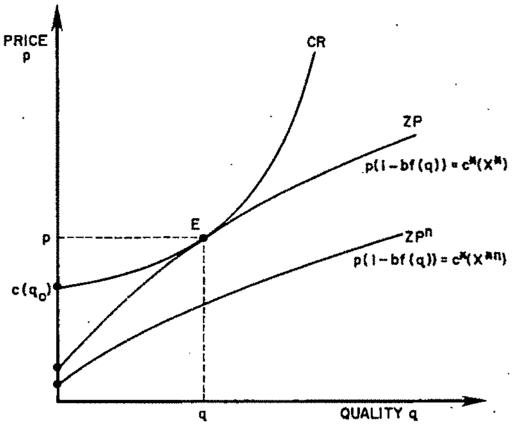


FIGURE 2. COUNTERFEITING EQUILIBRIUM

Potential counterfeiters earn zero profits when their unit production costs are equal to the "expected price," that is, the price of the goods for which the copies are being passed off times the probability that the fakes escape detection at the border. This gives the *zero-profit condition*,

$$p(1-bf(q))=c^*(X^*), \tag{8}$$

depicted as *ZP* in Figure 2.

In an equilibrium with $s \geq 0$, competition ensures that the zero-profit curve is tangent to the credibility constraint, and that the former curve is more concave than the latter.[12] In Figure 2, we show this tangency; as drawn, there are no points on or above the credibility curve that yield negative profits for counterfeiters. Algebraically, the *competition equation* requires

$$(1+r)c'(q)=\frac{bf'(q)p}{1-bf(q)}, \tag{9}$$

---

[11] As before, the deviant firm would not want to expand by "too much," or else consumers would (rightly) expect its rivals to respond in future periods. If they were to respond, the firm could not persist in its high level of sales. But then consumers would suspect quality shading in the current period.

[12] If at all quality levels the *ZP* curve is flatter than the *CR* curve, no such tangency will exist. Such is the case, for example, if $c^{*\prime}$ is very small. In this case, the market for high-quality goods collapses. As an example of this phenomenon, we cite the exit of Louis Vuitton from the Italian handbag market, following a period of intense competition with counterfeiters. See Kaikati and LaGarce (p. 58).

with $c''/c' > f''/f' + 2f'/(1-bf)$. Equations (7) and (9) jointly determine p and q; then (8) determines X^*.

We proceed to describe the remaining equilibrium conditions, still maintaining the assumption that these will be consistent with the existence of a positive amount of counterfeit-product trade. The break-even condition for legitimate, foreign producers of generic jeans is

$$p_0 = c^*(X^*). \tag{10}$$

The supply of foreign jeans must equal demand, where the latter includes both the demand by non-quality-conscious consumers and the unwitting purchases of counterfeits by the quality-conscious consumers. Sales of honestly labeled generic jeans account for N_0 pairs. Production of counterfeits must be $sN/(1-bf(q))$ if sN pairs are to survive border inspection. We assume at this point that customs agents discard the jeans that they confiscate. Thus, the market share of counterfeits is given implicitly by

$$X^* = N_0 + \frac{sN}{1-bf(q)}. \tag{11}$$

Finally, we record home firms' profits. In equilibrium, each trademark owner sells $(1-s)N/M$ pairs of jeans; per firm profits are

$$\pi = [p - c(q)](1-s)N/M - rF. \tag{12}$$

When entry is free, $\pi = 0$, which, together with (12), determines M.

We now are prepared to investigate whether trade in counterfeit products does in fact occur in equilibrium. Define π^{*n} as the marginal profitability of producing a copy when no counterfeiting actually takes place, that is, $\pi^{*n} \equiv p(1-bf(q)) - c^*(N_0)$. Let ZP^n in Figures 2, 3A, and 3B represent the points where $\pi^{*n} = 0$. In Figure 2, this curve does not intersect CR. In this case, it is profitable for some counterfeiters to enter, if none are present already, at *any* price-quality pair satisfying the credibility constraint. The equilibrium is at E, where the zero-profit condition associated with the X^* from (11) is tangent to CR. Since $X^* > N_0$, equilibrium has positive counterfeiting.

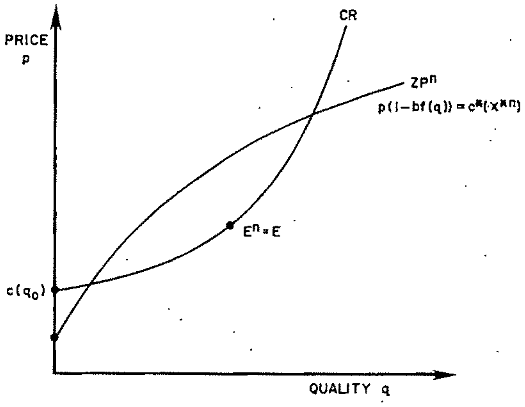


FIGURE 3A. COSTLY COUNTERFEITING ($s = 0$)

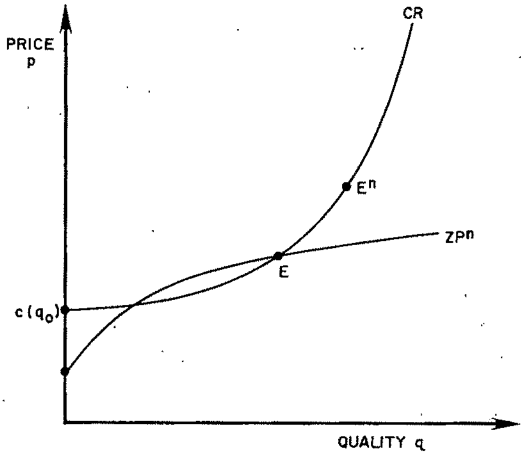


FIGURE 3B. LIMIT PRICING EQUILIBRIUM ($s = 0$)

If ZP^n intersects CR, as in Figures 3A and 3B, no counterfeiting takes place in equilibrium. These two figures represent qualitatively different situations. In Figure 3A, ZP^n passes above E^n, the equilibrium point for the no-counterfeiting regime. Home firms can select quality, price, and output as if counterfeiting were infeasible, and in the resulting equilibrium counterfeiters have no incentive to enter. The situation is different in Figure 3B. Here, production at (p^n, q^n) would invite entry by counterfeiters. Equilibrium occurs instead at the intersection of ZP^n and CR that is closer to the noncounterfeiting equilibrium. This may be to the right of point E^n (if both intersections of ZP^n with CR lie to the northeast of E^n) or to the left of point E^n (as drawn). Although counterfeiters make no sales in such a *limit-pricing equilibrium*, the threat of their entry affects the market outcome.

When $s > 0$, the output of generic jeans is increased by counterfeiting. Consequently, foreign costs rise with counterfeiting, and so does the price of low-quality jeans. The foreign country gains from counterfeiting, as its terms of trade improve. Home consumers of low-quality jeans qua consumers (i.e., neglecting their role as owners of firms) must lose.

Home producers may provide either higher- or lower-quality products in an equilibrium with (actual) counterfeiting than in one where counterfeiting is infeasible. Indeed, from (7) and (9) we see that θ no longer influences the equilibrium quality when $s > 0$. Although the home firms would like to compete for customers by further tailoring their products to consumers' tastes, they are constrained in doing so by the potential response of counterfeiters. Interestingly, the home firms may earn higher profits in an equilibrium with counterfeiting (assuming that M is fixed) than they would if it were infeasible. Price markups rise as we move up the credibility constraint,[13] so if quality increases due to counterfeiting, so too do per-unit profits. Then, if sales do not fall by too much, total profits rise as well. When profits increase, it is because the presence of counterfeiters limits the intensity of mutually harmful competition among the legitimate firms.

In a limit-pricing equilibrium, foreign output and sales of low-quality jeans are the same as in the no-counterfeiting regime. So too are foreign costs. Thus, the mere threat of counterfeiting has no effect on the foreign country, or on home-country consumers of low-quality jeans. Home producers select (credible) prices and qualities to maximize quality-conscious consumers' utilities, subject to the constraint that entry by counterfeiters be effectively deterred. The resulting level of quality can be higher or lower than in the no-counterfeiting equilibrium. For fixed M, profits per firm are higher in a limit-pricing equilibrium than in the no-counterfeiting equilibrium if and only if quality is higher.

C. *Determinants of the Market Share of Counterfeits*

In this subsection, we explore some of the determinants of the market share of counterfeits. We study here the influences on s of: the lag in information transmission, T; the discount factor, i; the shape of the home cost function, $c(q)$; and the (absolute and relative) sizes of the two market segments, N and N_0. As a by-product of the analysis, we also learn what factors affect market prices and quality levels.

Our approach is to calculate comparative-static derivatives of the system of equations (7) through (11). By doing so, we restrict attention to parameter values that imply an equilibrium with a positive amount of counterfeit-product trade. Note, however, that the factors that cause s to decline are also the ones that make it more likely that no counterfeiting will take place in equilibrium. To conserve space, we omit our algebraic computations.

The speed of information transmission T and the discount factor i both enter our model through the composite parameter r. A change in either of these primitive parameters has a similar effect on the incentive firms have to run down their reputations. For example, if T increases, consumers observe quality with a longer lag, and thus quality shading becomes more attractive. Similarly, an increase in i means that the future profits from maintaining a reputation are discounted more heavily, and again it is more tempting to cut quality for immediate gain.

When r increases, home firms find it more difficult to convince consumers that their goods are of high quality. In terms of Figure 2, the CR shifts up and becomes steeper. The ZP curve must likewise shift up to reestablish the tangency between the two. The new equilibrium is characterized by lower q and higher X^* and s. Quality of brand-name products falls, because the temptation to cheat is greater. The fall in quality reduces the risk to counterfeiters of confiscation, which in turn causes their market share to increase.

---

[13] Along CR, $d[p - c(q)]/dq = rc'(q) > 0$.

The shape of the home cost schedule reflects the marginal cost of quality. We can parameterize this shape (as distinct from the *level* of costs) by writing $c(q,\alpha)$, where $\partial c/\partial\alpha=0$ at the equilibrium q, and $\partial^2 c/\partial q\,\partial\alpha>0$. A higher value of α corresponds to a steeper $c(q)$ schedule. We investigate the comparative statics with respect to α, to understand the role played by the shape of the cost curve in determining the equilibrium share of counterfeits.

An increase in α causes the CR curve to shift up (at least near the initial equilibrium). When the marginal cost pivots about q, $c(q_0)$ falls. This means that home firms have a greater temptation to cheat. The restoration of equilibrium requires higher-foreign costs, hence additional output by counterfeiters. The new equilibrium has smaller q and larger s.

Finally, we consider the effects of shifts in demand. A change in N or in N_0 has no effect on the p and q determined by equations (7) and (9). Thus, by (8), X^* does not change. If N increases with N_0 constant, the total number of forgeries produced remains the same, as does the probability of confiscation. So the market share of counterfeits falls. If N_0 rises, given N, more of the (given) foreign output is sold honestly, and again s falls. An increase in the fraction of quality-conscious consumers, holding constant the total population, causes the market share of counterfeits to rise if and only if $1-\phi>s$. This condition is satisfied in all plausible cases.

We summarize our findings in

PROPOSITION 1: *The market share of counterfeits increases with*: (*i*) *increases in the quality-detection lag*; (*ii*) *increases in the marginal cost of quality*; (*iii*) *decreases in the size of the home market*; (*iv*) *increases in the fraction of quality-conscious consumers, if* $1-\phi>s$.

D. *Welfare Effects of Potential Counterfeiting*

We compare levels of social welfare in the no-counterfeiting and cum-counterfeiting equilibria. Our main result, proved in the Appendix, is

PROPOSITION 2: *With free entry by home firms, actual or potential counterfeit-product trade entails a loss in home and world welfare.*

Under conditions of free entry, the effect of potential or actual counterfeit-product trade on home-country welfare can be expressed as the sum of three terms. The first effect is the terms-of-trade loss suffered by non-quality-conscious home consumers. This term is nonzero whenever $s>0$. The second term is the direct loss to quality-conscious consumers that results from their being deceived. They buy counterfeit goods that have a market value of p_0 but pay p instead. This term also is strictly negative when $s>0$. The final term represents a "quality-adjustment effect." Because the quality level in the no-counterfeiting equilibrium with free entry is a constrained optimum, any change in that level induced by actual or potential counterfeiting entails a loss of welfare.

The first of these effects represents a transfer between home-country consumers and foreign producers. From the point of view of world welfare,[14] this effect washes out. When home customs agents discard the goods that they confiscate, the second effect does reflect a loss in world welfare, namely the opportunity cost of the counterfeits that are destroyed. This loss would not arise if customs agents were to auction off the goods that they seized.[15] Finally, the quality-adjustment effect also reflects a loss in world efficiency. So world welfare, like domestic welfare, falls.

---

[14] We study "world welfare" to understand the implications of counterfeiting for global efficiency. Of course, this concept has normative content only if international compensations are possible.

[15] We consider the implications for home-country welfare of alternative means of disposing the confiscated forgeries in Section III, Part B, below.

Surprisingly, counterfeit-product trade need not bring about a loss in domestic or world welfare when the number of home-owned, jeans-producing firms is fixed. For given M and absent counterfeting, the nationally optimal quality is the first best, where $\theta = c'(q)$. As we have noted, quality in the no-counterfeiting equilibrium falls short of this first-best level. For standard, second-best reasons, actual or potential counterfeiting can raise home welfare, if it causes a higher level of quality to obtain, that is, if $q > q^n$. The threat of entry by counterfeiters may remove the temptation firms otherwise have to compete via (credible) reductions in price and quality. Such competition, while beneficial to consumers, reduces home welfare.

Counterfeit-product trade is most likely to raise home welfare when the actual share of counterfeits is small, since the occurrence of $s > 0$ continues to imply a terms-of-trade loss for nonquality-conscious consumers and a direct surplus loss for quality-conscious consumers when the number of home firms is fixed. A gain in welfare also requires that $bf'(q)$ be relatively large (i.e., that home producers find that quality upgrading is an effective means of raising counterfeiters' costs of doing business).

Finally, we note that when M is fixed, increases in quality up to the first-best level reflect gains in worldwide efficiency. Thus, counterfeit-product trade can also raise global welfare in this case (and must do so whenever it raises domestic welfare). Indeed, if home-country customs agents do not discard the goods that they confiscate, world-welfare improvement can be consistent with quite high volumes of trade in counterfeit goods. In summary,

PROPOSITION 3: *With a fixed number of home firms, actual or potential counterfeit-product trade may raise or lower home or world welfare. Counterfeiting is more likely to raise welfare if quality upgrading deters counterfeiters effectively* (*i.e., if bf' is large*).

III. Policy Analysis

Only recently have governments in the more-developed countries begun to formulate policy responses to foreign counterfeiting. In the United States, for example, the trademark-protection and trade laws were amended in 1984 to allow for more severe punishment of offenders, and to stipulate that tariff preferences be denied to countries that harbor counterfeiters. Even now, a number of the key policy questions concerning counterfeiting remain unresolved.

Our model allows us to address two of the issues that have featured prominently in the public debate. The first is whether the government ought to devote more resources to searching for counterfeits at the border, as argued by many trademark owners who are the targets of forgeries. Despite these arguments, we have witnessed recently in the United States a trend reduction in U.S. customs enforcement. The second issue concerns the disposition of counterfeit goods seized at the border. Current U.S. law gives the Customs Service several options in this regard, and Congress has not as yet stipulated a procedure choosing among these.

A. *Border Inspection Policy*

We study in this section the efficacy of tighter border policy. In our model, we capture such a policy change by increases in the parameter b. Recall that b is the probability that a given shipment is examined by customs officials. Here we evaluate the *benefits* of stricter enforcement. Of course, a complete analysis would require that these benefits be weighed against the associated real resource costs.

We seek to determine dW/db, where $W \equiv N_0U_0 + NU$ is aggregate home-country welfare. We begin with the case of free entry. Then profits from local jean production are zero, and the income terms that enter U_0 and U can be taken as fixed. Differentiation of W with respect to b gives

$$\frac{dW}{db} = -N_0\frac{dp_0}{db} - N\theta(q-q_0)\frac{ds}{db} - sN\frac{dp}{db} + N(1-s)\left[\theta\frac{dq}{db} - \frac{dp}{db}\right]. \tag{13}$$

The four terms on the right-hand side of (13)

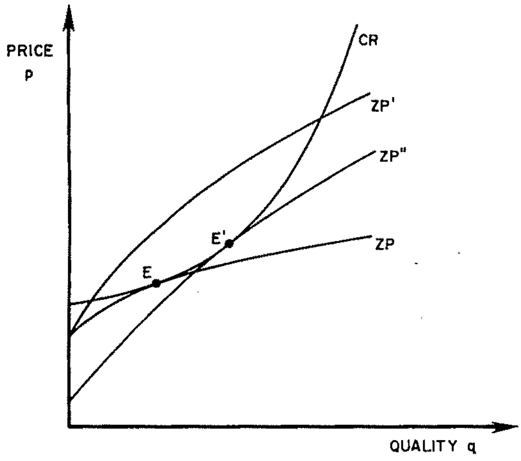


FIGURE 4. TIGHTER BORDER INSPECTION POLICY

are readily understood. The first is a terms-of-trade effect, reflecting any change in the price of imported low-quality jeans. The second term represents the direct effect on consumers' utilities of there being a different likelihood that a particular purchase will turn out to be fake. The third term reflects changes in the terms of trade on counterfeits. Finally, there is a quality-adjustment effect.

To sign these various effects, we need to know how a change in b alters the equilibrium. Throughout this section (and the next) we focus on equilibria that have $s > 0$. The border-policy parameter appears in equations (8) and (9). An increase in b raises the cost of delivering counterfeit products. At the same time, stricter enforcement implies a greater effectiveness of quality upgrading as a means of deterring counterfeiters. As shown in Figure 4, the ZP curve shifts up to ZP' and becomes steeper. Equilibrium is restored by a decline in foreign costs (effected by a reduction in X^*), which shifts the ZP curve down to ZP''. The new equilibrium E' has a smaller s and a higher q, the latter reflecting the optimal response of home producers to the policy change.

Two benefits of an increase in b are fairly obvious. First, quality-conscious consumers buy fewer fakes. Second, the shift of foreign resources out of counterfeit production causes p_0 to fall, thereby improving the home country's terms of trade. However, reflection on the last two terms in (13) reveals that the analysis is more subtle than one might initially suppose. The third term is unambiguously negative. An expansion of enforcement efforts causes the price of legitimate high-quality jeans to rise (along with q). But this implies that when quality-conscious consumers now are deceived by counterfeiters, they pay more for their mistakes. The last, quality-adjustment term can have either sign. Substituting for $dp/db = (1+r)c'(q)\,dq/db$, we see that this term is positive exactly when the equilibrium quality of home jeans is below that of the no-counterfeiting equilibrium.

Combining these four effects, we find that an increase in b need not be welfare improving, even if we neglect the resource expenditure needed for this policy measure.[16] The explanation lies in the "general theory of the second best": policies with direct effects that are beneficial can nonetheless have deleterious consequences if their indirect effect is to exacerbate a preexisting market distortion. An increase in b, which causes q to rise when $s > 0$, can move the equilibrium level of quality away from the (constrained) optimum, if this level initially is too high. We summarize in

PROPOSITION 4: *With free entry by home firms and* $s > 0$, *tighter enforcement of border policy can raise or lower home welfare. A sufficient condition for* $dW/db > 0$ *is* $c'(q) < \theta(1-s)/(1+r)$.[17]

We turn briefly to the case of fixed M. Then the home industry earns positive profits, which enter into consumers' utilities through the income variables, y and y_0. The new expression for dW/db contains one new term and a somewhat different quality-adjustment effect. The other (three) terms are exactly as before.

---

[16] In a limit-pricing equilibrium (with free entry), an increase in b always raises welfare. The first three terms of (14) all vanish in this case, while relaxation of the threat of entry by counterfeiters allows the legitimate firms to offer greater surplus to consumers. This reconciles our result here that dW/db might be negative with our earlier finding that welfare is always higher when $\phi \equiv 1$; the increase in W need not be monotonic.

[17] When the government sells confiscated goods rather than destroying them, a sufficient condition for $dW/db > 0$ is $c'(q) < \theta/(1+r)$, that is, $q < q^n$.

The new term reflects the increase in the profits of home firms that results from their capturing sales otherwise made by counterfeiters. It is given by $-N(p-c(q))\,ds/db$. The quality-adjustment effect that applies when M is fixed is $N(1-s)[\theta-c'(q)]\,dq/db$. This term is positive exactly when the equilibrium quality of legitimate jeans falls short of the first-best level.

Once again, the total effect on home welfare of an increase in b is ambiguous. Combining the quality-adjustment term and the terms-of-trade effect on imported counterfeits, we find that a sufficient condition for $dW/db>0$ when M is fixed is $c'(q)<\theta(1-s)/(1+rs)$. This condition is less stringent than in the free-entry case, because tighter border policy in the present circumstances does not cause socially wasteful entry.

B. *Government Disposition of Confiscated Merchandise*

We consider now the issue of the disposition of goods seized by customs officials. To this point, we have assumed that all confiscated counterfeits are destroyed. But in fact, U.S. law gives the Customs Service four options in this regard. They may allow reexportation of the goods, donate them to charity, destroy them, or turn them over to the General Services Administration (GSA) for relabeling and sale. In practice, most confiscated merchandise is turned over to the GSA, which relinquishes it to a charity if a request is received within a year. After that period, the goods are sold at auction. Of course, counterfeits deemed to be dangerous are destroyed.[18]

To study what policy of disposition is optimal, we introduce into our model a parameter σ, which represents the fraction of confiscated jeans that are resold. We assume that the government removes the bogus trademark from these goods and offers them at the competitive price p_0.

How does this policy variable affect the equilibrium? When $\sigma>0$, the home government becomes an additional supplier in the market for low-quality jeans. The condition for equilibrium in this submarket changes to

$$X^*+\sigma bf(q)\frac{sN}{1-bf(q)} = N_0+\frac{sN}{1-bf(q)}, \tag{11'}$$

where $sN/(1-bf(q))$ is the number of counterfeits produced and $\sigma bf(q)$ is the fraction of these that are seized and auctioned by the home government. The remaining conditions of equilibrium are unchanged. Importantly, this means that σ does not affect the equilibrium values of p, q, or X^*, which continue to be determined by (7), (8), and (9).

Now consider the welfare implications of variations in σ. Such changes have only two effects in the model. First, as σ is increased, the government collects additional revenue of $[p_0bf(q)sN/(1-bf(q))]\,d\sigma$. This revenue represents the direct benefit from reducing waste. But, the increase in σ also causes the market share of counterfeits to increase. Since X^* is constant, each pair of low-quality jeans sold by the home government, by crowding out one pair of generic imports, causes one more pair to be falsely labeled. In effect, by entering the market as a supplier of low-quality jeans, the home government forces foreign producers out of legitimate activities and into counterfeiting. The welfare cost associated with this increase in s under conditions of free entry by home firms is

$$\left\{\theta(q-q_0)-\frac{\sigma p_0}{1-\sigma bf(q)}\right\}\frac{sNbf(q)}{1-bf(q)}\,d\sigma.$$

The first term in braces captures the loss to quality-conscious consumers, while the second reflects the gain in government revenue. Each term is multiplied by $sNbf(q)/(1-bf(q))$, the total number of units confiscated.

---

[18] Testimony of Richard H. Abbey, Chief Counsel for the U.S. Customs Service, before the House Subcommittee on Crime. See U.S. House of Representatives Committee on the Judiciary (1985, p. 138).

Adding the two effects together, and rearranging terms, we find

$$\frac{dW}{d\sigma}=\frac{sNbf(q)}{1-\sigma bf(q)}\left[p-\theta(q-q_0)\right]. \tag{14}$$

Expression (14) can be positive or negative. But notice that its sign is independent of σ. This proves

PROPOSITION 4: *With free entry by home firms, the optimal policy is to sell (discard) all confiscated counterfeits when* $p>(<)\theta(q-q_0)$.

Selling the confiscated items is most likely to be socially beneficial when the difference in quality between domestic and foreign jeans is small.[19]

IV. Conclusions

When foreign firms forge the trademarks of home manufacturers, they infringe on the property rights of the legitimate producers. Counterfeiting can only be profitable when brand-identifying trademarks, and the reputations that they embody, have significant value to their owners. For reputations to be of value, consumers must be imperfectly informed about product attributes. So, counterfeiting must be studied in the context of imperfect consumer information.

We conduct our analysis of counterfeit-product trade in a dynamic, two-country model with imperfect quality information and brand-name reputations. We solve the model for a steady-state, rational expectations equilibrium, and use it to study the causes and consequences of counterfeits. The confluence of imperfect information and imperfect property rights gives a second-best flavor to our welfare analysis.

The presence of foreign counterfeits harms the home economy in a number of intuitive ways. Some home consumers suffer when they unwittingly purchase copies. And the home country realizes a deterioration of its terms of trade, as foreign counterfeit production drives up factor prices in the export sector of the foreign country.

The feasibility of counterfeiting also alters the rivalry among domestic manufacturers. Absent counterfeiting, brand-name producers compete to offer consumers the best deal, subject to a credibility constraint that each firm find it optimal to provide its reputed quality. Potential counterfeiting forces firms to adjust their price and quality so as to protect themselves from imitators, while still competing for consumers. Depending upon the border inspection technology, the possibility of counterfeiting may raise or lower equilibrium quality and price. With free entry by home firms, the quality adjustment in response to counterfeiting necessarily lowers home and global welfare. With a fixed number of home firms, however, brand-name producers may raise their quality in an effort to battle counterfeiters, and this quality enhancement may cause both home and global welfare to rise (since quality initially was undersupplied due to imperfect information).

Policy responses to counterfeiting may have unintended consequences due to the changes they induce in the quality of brand-name products. Tighter border inspections policy, for example, benefits the home country in that it reduces the price of legitimate imports and lowers the market share of counterfeits. It may nonetheless lower domestic welfare as it causes the quality of brand-name products to increase (perhaps excessively) and worsens the terms of trade on the remaining counterfeits.

Another policy question concerns the disposition of counterfeit products that are confiscated at the border. We consider a policy whereby the customs authority sells a fraction σ of these items and destroys the rest. We find that increasing the fraction that are sold actually raises the market share of counterfeits. The home government's supply of low-quality goods competes with that of legitimate foreign producers, and thereby shifts resources abroad into the illegal subsector. We show that the optimal policy is either to destroy all confiscated products or to auction them all, and we provide a simple sufficient condition that determines which of these is the case.

---

[19] Qualitatively, the same conclusion holds when the number of home firms is fixed. Then sale rather than destruction is indicated if and only if $c(q)>\theta(q-q_0)$.

APPENDIX: PROOF OF PROPOSITION 2

Home welfare is given by $W \equiv N_0 U_0 + NU$. When entry is free, profits are zero, so $y^n = y$ and $y_0^n = y_0$. Therefore,

$$
\begin{aligned}
W - W^n = & - N_0 (p_0 - p_0^n) \\
& + N[(1-s)\theta(q - q^n) \\
& + \theta s(q_0 - q^n) - (p - p^n)].
\end{aligned}
$$

Using the credibility constraint to substitute for p^n and p, we find

$$
\begin{aligned}
\text{(A1)}\quad & W - W^n \\
& = N_0(p_0^n - p_0) + Ns\theta(q_0 - q) \\
& + N\{\theta(q - q^n) - (1+r) \\
& \times [c(q) - c(q^n)]\}.
\end{aligned}
$$

The first term in (A1) is nonpositive. Since, $X^{*n} \leq X^*$, home consumers of low-quality goods pay no more for their jeans in the no-counterfeiting regime, and pay strictly less if $s > 0$ with counterfeiting. The second term is the direct loss to quality-conscious consumers from the presence of counterfeits in the market. The final term is the indirect effect on the surplus enjoyed by quality-conscious consumers caused by the induced change in quality. This too is nonpositive, because $c'' > 0$ implies that $[c(q) - c(q^n)] > c'(q^n)(q - q^n)$, so

$$
\begin{aligned}
\text{(A2)}\quad & N\{\theta(q - q^n) \\
& - (1+r)[c(q) - c(q^n)]\} \\
& < N(q - q^n)\{\theta - (1+r)c'(q^n)\} = 0.
\end{aligned}
$$

Let total revenue from production in the foreign country, excluding the value of jeans output, be $R^*(X^*)$, with $R^{*\prime} < 0$ and $R^{*\prime\prime} < 0$. The change in global welfare is $\Delta G \equiv (W - W^n) + (W^* - W^{*n})$. Substituting for $W^* - W^{*n} = R^*(X^*) - R^*(X^{*n}) + p_0 X^* - p_0^n X^{*n}$ and noting (11), we have

$$
\begin{aligned}
\Delta G = & R^*(X^*) - R^*(X^{*n}) + Ns\theta(q_0 - q) \\
& + N[\theta(q - q^n) - (p - p^n)] \\
& \times + sNp_0/(1 - bf(q)).
\end{aligned}
$$

But $R^{*\prime\prime} < 0$ implies $R^*(X^*) - R^*(X^{*n}) < R^{*\prime}(X^{*n})(X^* - X^{*n})$. Using $R^{*\prime}(X^{*n}) = -p_0^n$, we have $R^*(X^*) - R^*(X^{*n}) < -sNp_0^n/(1 - bf(q))$. Therefore,

$$
\begin{aligned}
\Delta G/N < & \theta(q - q^n) - (p - p^n) + s\theta(q_0 - q) \\
& + s(p_0 - p_0^n)/(1 - bf(q).
\end{aligned}
$$

The fact that quality-conscious consumers choose to consume the high-quality jeans in the no-counterfeiting equilibrium implies $p_0^n > \theta(q_0 - q^n) + p^n$. This, together with $p = p_0/(1 - bf(q))$, implies

$$
\begin{aligned}
\text{(A3)}\quad \Delta G/N < & (1-s)[\theta(q - q^n) - (p - p^n)] \\
& - sbf(q)p_0^n(1 - bf(q)) \\
= & (1-s)[\theta(q - q^n) - (1+r) \\
& \times (c(q) - c(q^n))] \\
& - sbf(q)p_0^n/(1 - bf(q)).
\end{aligned}
$$

The first term on the right-hand side of (A3) is a "quality-adjustment effect." We have already shown this to be nonpositive in (A2). The second term is the value (at no-counterfeiting prices) of the output discarded by home customs agents. This term would vanish if customs agents were to relabel and sell at auction the jeans they confiscate. In any event, $\Delta G < 0$.

REFERENCES

Allen, Franklin, "Reputation and Product Quality," *Rand Journal of Economics*, Summer 1984, *15*, 311–27.

Grossman, Gene M. and Shapiro, Carl, "Foreign Counterfeiting of Status Goods," *Quarterly Journal of Economics*, forthcoming 1988.

Kaikati, Jack G. and LaGarce, Raymond, "Beware of International Piracy," *Harvard Business Review*, March-April 1980, *58*, 52–60.

Klein, Benjamin and Leffler, Keith B., "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, August 1981, *89*, 615–41.

Olenick, Shari D., "Draft International Anticounterfeiting Code: Neo-Realism as a Vehicle for Analyzing the Effect of Nonsignatories' Perceptions on the Development of an Anticounterfeiting Norm," *Vanderbilt Journal of Transnational Law*, Fall 1982, *15*, 803–62.

Salmans, Sandra, "Locking Out the Product Poachers," *International Management*, June 1979, *34*, 14–19.

Shapiro, Carl, "Consumer Information, Product Quality, and Seller Reputation," *Bell Journal of Economics*, Spring 1982, *13*, 20–35.

______, "Premiums for High Quality Products as Returns to Reputations," *Quarterly Journal of Economics*, November 1983, *98*, 659–80.

Business Week, "The Counterfeit Trade: Illegal Copies Threaten Most Industries—and Can Endanger Consumers," December 16, 1985, 64–72.

New York Times, "Trade Officials to Push Counterfeit Goods Code," January 20, 1986.

U.S. House of Representatives Committee on Energy and Commerce (1983),"Unfair Foreign Trade Practices: Part I," Hearings before the Subcommittee on Oversight and Investigation, Serial No. 98-74.

______ (1984),"Unfair Foreign Trade Practices: Part III," Hearings before the Subcommittee on Oversight and Investigation, Serial No. 98-164.

U.S. House of Representatives Committee on the Judiciary (1985),"Trademark Counterfeiting Act of 1984," Hearings before the Subcommittee on Crime, Serial No. 151.

U.S. International Trade Commission, *The Effects of Foreign Product Counterfeiting on U.S. Industry*, 1984, USITC Publication No. 1479.

Foreign-Owned Land

By JONATHAN EATON*

Foreign investment affects land prices as well as domestic capital. A permanent increase in net foreign investment can reduce steady-state welfare if a consequence is higher land values. Other things equal, more crowded countries have higher land prices and lower permanent welfare, but their net foreign indebtedness depends on technology, savings behavior, and the interest rate. Even when the domestic capital stock is not affected, a land tax raises steady-state welfare.

Foreign investment contributes to the productive resources of the capital-importing country. It also affects the value of domestic assets in fixed supply. Most economic analysis, in assuming that assets in the host country are completely reproducible, addresses only the first effect. If some assets in the economy are supplied inelastically, however, the effect of foreign investment on asset values can reverse many standard results on the consequences of foreign investment for welfare, and a number of other comparative static results as well.

This paper develops a life-cycle model of an economy in which three factors, called land, labor, and capital, produce output. Land and capital serve as stores of value as well. The fundamental distinction between them is that capital is reproducible whereas land is in fixed supply.

The distinction between land and capital both as factors of production and as assets appears in David Ricardo (1817) and more explicitly in Henry George (1898). For George, the fundamental difference is that the productive contribution of land, unlike that of labor or capital, does not derive directly or indirectly from human endeavor, its role in production being "passive." George argues that private ownership of land diverts resources away from more productive activities:

> ..."It is seen that private property in land, instead of being necessary to its improvement and use, stands in the way of improvement and use, and entails an enormous waste of productive forces;" [1904, p. x].

In the analysis here, ownership of land, or more specifically private trade in claims on income from land, plays exactly this role by diverting savings away from investment in capital.

Raymond Goldsmith (1985) provides data that give some indication of the quantitative importance of land holdings. Table 1 presents his data on land as a share of tangible assets for twenty countries. These suggest that land can be a nontrivial component of wealth.[1]

One purpose of the analysis here is to consider the consequence of an exogenous increase in the level of foreign investment where land and capital both serve as factors of production and as a store of value. Peter Diamond (1965) and John Kareken and Neil Wallace (1977) address this issue in special cases of the model developed here.[2] The first

---

*Department of Economics, University of Virginia, Charlottesville, VA 22901, National Bureau of Economic Research, and Hoover Institution. I thank an anonymous referee for useful comments and the National Science Foundation for support under grant no. SES-8410613.

[1] Elsewhere I have explored the implications of changes in the terms of trade and in factor supplies in a three-factor, life-cycle model with portfolio autarky; that is, in the absence of foreign investment (Jonathan Eaton, 1987). Jean Tirole (1985) discusses the implications of fixed factors for asset bubbles. Oded Galor (1986) analyzes migration in an overlapping generations model.

[2] Willem Buiter (1981) extends Diamond's analysis to two countries.

TABLE 1—LAND AS PERCENT OF TANGIBLE ASSETS

	Approximate Year					
	1850		1913		1978	
Country	All Land	Agricultural Land	All Land	Agricultural Land	All Land	Agricultural Land
Australia	–	–	–	–	12.5	–
Belgium	54.5	–	18.7	–	18.1	–
Canada	–	–	–	–	24.3	3.4
Denmark	–	–	26.3	13.3	14.2	2.9
France	56.5	49.9	33.2	25.4	23.4	7.9
Germany	45.8	43.2	26.9	20.2	24.9	12.6
Great Britain	29.9	19.9	22.2	9.9	14.9	2.5
Hungary	–	–	–	–	27.1	16.5
India	35.3	–	53.7	–	25.1	–
Israel	–	–	–	–	18.5	3.3
Italy	47.2	41.8	39.2	33.3	16.8	5.1
Japan	–	–	42.2	29.7	50.7	11.0
Mexico	–	–	–	–	19.3	5.8
Norway	–	–	31.3	22.4	15.8	5.3
Russia/USSR	–	–	61.5	–	22.8	–
South Africa	–	–	28.0	21.1	18.5	8.5
Sweden	–	–	–	–	15.3	2.8
Switzerland	–	–	36.0	21.7	17.2	2.5
United States	41.3	35.8	35.3	19.2	24.9	5.4
Yugoslavia	–	–	–	–	18.5	15.4

Source: Goldsmith (1985, Tables 39 and 40).

analysis excludes land and the second capital. The two yield diametrically opposite conclusions about the implications of foreign investment for welfare, both on impact and in steady state. With no land, foreign capital inflow lowers the welfare of the currently retired but, if the equilibrium is efficient, steady-state welfare rises. With no capital, these results are just reversed. With both factors present the effects, not surprisingly, are ambiguous, but under specific assumptions about technology simple conditions determine their direction. Inserting data from several countries into these conditions suggests no presumption that a reaction in one direction or the other is universally more likely.

A second concern of this analysis is the behavior of a national economy when capital is perfectly mobile, so that the level of foreign investment equates domestic and world rates of return. An exogenous change in the world interest rate has an ambiguous effect on the value of land and on welfare. A fall in this rate can, for example, reduce the steady-state welfare even of a capital importer if it causes land prices to rise, while, on impact, the retired generation benefits from the consequent capital gain on land. An increase in the ratio of labor to land necessarily raises land prices and reduces steady-state welfare, however, while, again, on impact the currently retired benefit from the resulting capital gain.

The third issue that the paper addresses is the effect of a land tax on foreign investment and on national welfare. This analysis extends to an open economy work by Martin Feldstein (1977), Guillermo Calvo et al., (1979) and Christophe Chamley and Brian Wright (1987) who show, in a closed economy, that taxing land raises steady-state welfare. Such a tax raises the capital stock, causing the wage rate to rise and the interest rate to fall.[3] With perfect capital mobility, the domestic capital stock and factor prices are unchanged as a consequence of a land tax. Nevertheless, a tax on land raises steady-state welfare as long as the value of land remains positive. Thus George's (1904) argument in favor of such a tax extends to an economy facing a given world interest rate, even though the total capital stock is not affected by the tax.

---

[3] The effect on the generation first subject to the tax is necessarily negative, however. This result typically extends to the open economy, as I discuss in Section IV.

Section I presents the basic assumptions of the analysis. The effect of an exogenous increase in foreign investment is examined in Section II. Section III analyzes the economy under perfect capital mobility, while the issue of the land tax is addressed in Section IV. Section V provides some concluding remarks.

I. The Model

The economy has a constant endowment of land, in an amount T, and a constant labor force L. Investment in the previous period determines the stock of capital in period t, denoted K_t. Aggregate output net of depreciation in period t, Q_t, is a homogeneous, twice-differentiable function of the three inputs,

$$Q_t = G(K_t, L, T). \tag{1}$$

Standard marginal productivity conditions determine the factor rewards w_t for labor, r_t for capital (net of depreciation), and π_t for land. I adopt the normalization $T=1$.[4]

New investment in capital takes the form of currently produced units of output that are not consumed. National savings is determined by a simple life-cycle optimization. Individuals live two periods. In the first period of life they provide a unit of labor services to earn a wage w. An amount c^y is spent on current consumption with the remainder invested in land and capital. In the second period the individual consumes the value of his assets and the income that they have earned. This amount is denoted c^0. Lifetime utility is $U(c^y, c^0)$.

The price of land is q_t. Denoting the individual's investment in capital as k_t and in land as l_t, budget constraints imply that

$$c_t^y = w_t - k_{t+1} - q_t l_{t+1}. \tag{2}$$

$$c_{t+1}^0 = (1 + r_{t+1})k_{t+1} + (q_{t+1} + \pi_{t+1})l_{t+1}. \tag{3}$$

With perfect foresight, for both land and capital to be held in positive amounts requires that

$$\frac{q_{t+1} + \pi_{t+1}}{q_t} = 1 + r_{t+1}, \tag{4}$$

which implies, with (3), that

$$c_{t+1}^0 = (1 + r_{t+1})(w_t - c_t^y). \tag{5}$$

In the working period the individual's problem is to choose c_t^y to maximize utility, given w_t and r_{t+1}. The utility-maximizing level is denoted by the function $c^y(w_t, r_{t+1})$.

Equilibrium in the markets for land and capital implies that

$$K_{t+1} = Lk_{t+1} + K_{t+1}^f \tag{6}$$

$$1 = Ll_{t+1} + T_{t+1}^f, \tag{7}$$

where K_{t+1}^f and T_{t+1}^f denote net foreign investment in capital and land, respectively. Negative values of these variables imply net ownership of these assets abroad by nationals.

The supply of capital and the price of land q_t evolve according to the two dynamic equations

$$K_{t+1} = L[w_t - c^y(w_t, r_{t+1})] - q_t + F_t \tag{8}$$

$$q_t = \frac{\pi_{t+1} + q_{t+1}}{1 + r_{t+1}}, \tag{9}$$

where F_t denotes net foreign investment in period t.

---

[4] This specification need not imply that land is homogeneous or that all land is actually used in production. See Eaton (1987).

II. Foreign Investment

This section treats net foreign investment as exogenous, and considers the effect of changes in its amount on the capital stock, the price of land and national welfare, both on steady-state configurations and in the transition from the initial situation to the new steady state.

A. *The Capital Stock and Price of Land*

In a steady state with a constant level of net foreign investment F, the capital stock and price of land are constant values, $\bar{K}$ and $\bar{q}$, respectively, that satisfy[5]

$$\bar{K} = \chi(\bar{K}) - \bar{q} + F \tag{10}$$

$$\bar{q} = \bar{\pi}/\bar{r}, \tag{11}$$

where

$$\chi(\bar{K}) \equiv L[\bar{w} - c^y(\bar{w}, \bar{r})],$$

total savings of workers in steady state, and

$$\bar{w} \equiv G_L(\bar{K}, L, 1)$$

$$\bar{r} \equiv G_K(\bar{K}, L, 1)$$

$$\bar{\pi} \equiv G_T(\bar{K}, L, 1).$$

An increase in total net foreign investment is indicated by an increase in F. Regardless of the form that investment takes, the effects on $\bar{K}$ and $\bar{q}$ are given by

$$\frac{d\bar{K}}{dF} = \frac{1}{\Delta} \tag{12}$$

$$\frac{d\bar{q}}{dF} = \frac{q_K}{\Delta}, \tag{13}$$

where

$$\Delta \equiv 1 - L(1 - c_w^y) w_K + L c_r^y r_K + q_K, \tag{14}$$

$$q_K \equiv \frac{\pi_K - q r_K}{r}. \tag{15}$$

As shown in the Appendix, as long as $1 + L c_r^y r_k \geq 0$ and $c_w^y \leq 1$, the condition $\Delta \geq 0$ is necessary and sufficient for the existence of a unique, stable, nonoscillating convergence path to steady state. This condition states that in the neighborhood of steady state an increase in the capital stock by a particular amount raises investment in capital by a lesser amount.

If this stability condition is satisfied, then an increase in foreign investment raises the steady-state capital stock. The interest rate falls and the return to at least one other factor must rise. If land and capital are complements in production, then the price of land necessarily rises, since $\pi_K > 0$.

An illuminating special case to consider is one in which the aggregate production technology derives from the specific-factors model described by Ronald Jones (1971) and Paul Samuelson (1971). One commodity, a manufactured good, uses only capital and labor in its production. An agricultural good uses only land and labor. Denoting the outputs of the manufactured good and agricultural good as Q^M and Q^A, respectively, and the respective labor inputs as L^M and L^A,

$$Q_t^M = G^M(K_t, L_t^M)$$

$$Q_t^A = G^A(T, L_t^A),$$

where G^M and G^A are homogeneous, twice-differentiable functions. Full employment ensures that $L_t^M + L_t^A = L$.

With the manufactured good serving as numeraire and denoting the relative price of the agricultural good as p_t, the value of output is $Q_t = Q_t^M + p_t Q_t^A$, which is maximized when value of the marginal product of labor is the same in the two sectors. If L_t^{M*} is the value of L_t^M that satisfies this condition then the aggregate production function, which now depends on p_t as well as aggregate factor supplies, is

---

[5] Conditions for the existence of a $\bar{K}$ and $\bar{q}$ that satisfy (10) and (11) are provided elsewhere (Eaton, 1987). A condition for the stability of the steady state is given in the Appendix.

$$(1') \qquad Q = G(K_t, L_t, T, p_t) = G^M(K_t, L_t^{M*}) + p_t G^A(T, L - L_t^{M*}).$$

In the specific-factors model an increase in foreign investment necessarily *lowers* the return to land. This is because the increase in the capital stock attracts labor to the manufacturing sector.[6] Since the interest rate discounting these returns also falls, the effect on the value of land remains ambiguous. It depends upon the relative labor shares in the two sectors. If the share of labor in manufacturing exceeds its share in agriculture, then the price of land rises. Otherwise it falls. An equivalent condition is that foreign investment raises the price of land if the share of land in domestic wealth, $q/(q+K)$, exceeds the ratio of the labor force in agriculture to total labor. Otherwise foreign investment lowers the price of land.

In the first case the value of land is primarily affected by its role as an asset. Hence it appreciates as a consequence of the lower interest rate. In the second, the value of land is primarily affected by its role as a factor of production complementary with labor. Its value then falls as a consequence of the higher wage.

To demonstrate these results, observe that in the specific-factors model, $\pi_K = -w_K(L - L_M)$ and $r_K K + w_K L_M = 0$. Hence

$$q_K = \frac{\pi w_K}{wr}\left[\frac{wL_M}{rK} - \frac{w(L-L_M)}{\pi}\right],$$

which is positive or negative as

$$\frac{wL_M}{rK} \gtrless \frac{w(L-L_M)}{\pi},$$

or, rearranging, as

$$\frac{q}{K+q} \gtrless \frac{L-L^M}{L}.$$

Thus there is no presumption that a permanent increase in foreign investment raises the price of land as it does in Kareken and Wallace's (1977) model. If agriculture is labor-intensive relative to manufacturing, the price falls.[7]

B. *Foreign Investment and Welfare*

The implications of a permanent increase in the level of foreign investment for the welfare of current and future generations can be very different. In the absence of a fixed factor, the case of Diamond's analysis, an increase in foreign investment necessarily lowers the welfare of the generation that enters retirement when the increase in foreign investment first occurs. It lowers the rate of return on that generation's savings while not contributing to their wage income in earlier years. Foreign investment can consequently not provide a Pareto improvement. If the steady state is efficient, so that the interest rate exceeds the population growth rate, later generations benefit. The positive effect of the increase in the wage more than offsets the negative effect of the lower return on savings. (In an inefficient equilibrium this is not the case: an increase in foreign investment lowers welfare of subsequent generations as well. It can then be Pareto worsening.)

---

[6] This result generalizes to technologies in which both sectors employ all three factors as long as land is the factor used most in agriculture and capital the factor used most in manufacturing. See Roy Ruffin (1984).

[7] Except for Hungary and Yugoslavia, in the most recent period for which Goldsmith provides data, agricultural land constitutes less than half of total wealth in land, as indicated in Table 1. The specific-factors model apparently does not apply. For Yugoslavia, however, nonagricultural land is insignificant, so that here the specific-factors model could have some relevance. The fraction of the labor force in agriculture in this country in 1978 was 33 percent, substantially in excess of the share of land in wealth (World Bank, 1980). If the specific-factors model is appropriate, then foreign investment acts to *lower* the price of land in this country.

A fixed asset eliminates the possibility that the steady state is inefficient, since the return on that asset must at least equal the population growth rate. It can then never be the case that an anticipated increase in foreign investment is Pareto worsening. It either raises steady-state welfare or it raises the welfare of the retiring generation upon announcement. It can even do both. When the increase in foreign investment is unanticipated, however, then the effect on the contemporaneous retired nationals depends, among other things, on how ownership of domestic land and capital is divided between nationals and foreigners.

I first consider the implications of foreign investment for the steady state. For one thing, this is the relevant criterion for comparing two parallel economies that have different, permanent levels of foreign investment. I discuss the transitional effects of an increase in foreign investment in this section, Part C.

Differentiating utility, U, evaluated at steady state, with respect to F gives, using the first-order condition for utility maximization,

$$\frac{dU}{dF}=\frac{\partial U}{\partial c_0}\left[r_K(w-c^y)+(1+r)w_K\right]\Delta^{-1}. \tag{16}$$

Incorporating the relationship

$$w-c^y=\frac{q+K-F}{L},$$

the Euler condition (that $\pi_K=-w_KL-r_KK$), and equation (15), this expression can be shown to have the sign of

$$r(w_KL-q_K)-r_KF, \tag{17}$$

if there is no national investment abroad.

The last term in this expression is positive. As in G. D. A. MacDougall's (1958) analysis of foreign investment, an exogenous increase in foreign investment acts to raise welfare by lowering the return on existing foreign-owned assets.

The first term can have either sign. The availability of land with a nonnegative yield as a store of value ensures that the steady state is efficient, that is, that $r\geq 0$. Consequently foreign investment raises steady-state welfare if it raises the real wage ($w_K>0$) and lowers the price of land ($q_K<0$). In Diamond's (1965) model with no land, w_K is necessarily positive, so that the outcome depends upon whether the initial equilibrium is efficient or inefficient. In Kareken and Wallace's (1977) model with no capital, $w_K=0$ while $q_K<0$, so that foreign investment reduces welfare in steady state.

Because of the generally ambiguous permanent welfare effect of foreign investment some special cases merit discussion: (*i*) If labor and capital are substitutes in production ($w_K\leq 0$), then necessarily $\pi_K\geq 0$, so that $q_K\geq 0$. Foreign investment necessarily *lowers* welfare in steady state if initially $F=0$. (*ii*) If the aggregate production function is Cobb-Douglas, so that

$$Q_t=K_t^\alpha L^\beta \qquad \alpha+\beta<1$$

then

$$q=\frac{1-\alpha-\beta}{\alpha}K$$

and

$$w_KL-q_K=\alpha\beta K^{\alpha-1}L^\beta-\frac{1-\alpha-\beta}{\alpha}$$
$$=r\beta-\frac{1-\alpha-\beta}{\alpha}.$$

Foreign investment acts to raise steady-state welfare if the interest rate is initially high and the share of land $(1-\alpha-\beta)$ is small relative to that of capital (α).[8] (*iii*) In the specific-factors model expression (17) has the sign of

---

[8] Edward Denison (1974) estimates the shares of earnings in the nonresidential business sector of the United States by input in 1968 as follows: Labor = 79.94, nonresidential structures and equipment and inventories = 16.02, and land = 4.04. Under a Cobb-Douglas assumption the expression is positive or negative depending upon whether $r \gtrless 5$ percent. If a period

$$\frac{(1+r)K}{L^M}-\frac{K+q-F}{L}.$$

A sufficient condition for foreign investment to raise steady-state welfare is that the capital-labor ratio in manufacturing exceed wealth per worker or, equivalently, that the ratio of the labor force in agriculture to total labor exceed the share of land in *national* wealth.

C. *Dynamic Adjustment*

The discussion so far has compared the effect of different levels of foreign investment across steady states. The dynamics of a transition from one steady state to another can be analyzed by manipulation of equations (A3) and (A4) in the Appendix. Once the change in foreign investment has occurred, the capital stock and price of land, and consequently factor prices, begin to move monotonically toward their new steady-state values.

If the change is anticipated then the land price begins to move toward its new steady-state value prior to the actual increase in foreign ownership. If the steady-state price of land falls, then the capital stock begins to rise before the increased foreign investment occurs. As the price begins to fall in anticipation, more national savings is diverted toward capital investment. If the steady-state land price rises, however, then there is a period of capital *decumulation* before the increase in foreign investment. Once the investment takes place, this process is reversed as the capital stock moves to exceed its initial steady-state level.

The older generation when the increase in foreign investment is first anticipated experiences a capital gain or loss on its land holdings depending upon whether the ultimate effect is to raise or to lower the steady-state land price. Since an increase in the steady-state price of land is *necessary* for foreign investment to lower steady-state welfare, anticipated foreign investment is never Pareto worsening.

What happens, on impact, in the case of an unanticipated increase in foreign investment depends on the initial distribution of assets between nationals and foreigners. If the land price rises and nationals hold relatively more land than capital, then the current welfare effect, from a national perspective, is more likely to be positive.

To summarize, in the one-asset Diamond (1965) model, foreign investment results in a capital deepening that raises the wage and lowers the interest rate. In that analysis, the retired generation when the investment occurs experiences a loss in welfare, but if the interest rate exceeds the growth rate then the long-run effect on welfare is positive. These results extend to an economy with land if there is no positive effect on the price of land. In this case current wealth holders experience a capital loss when foreign investment is first anticipated. Subsequent generations benefit from the higher wage.

If the price of land rises, then any wage increase may *not* compensate for the fall in the interest rate in determining the overall effect on steady-state welfare. But in this case current wealth holders benefit from a capital gain when foreign investment is first expected.

III. Perfect Capital Mobility

Section II considered the effect of an increase in foreign investment, treating this amount as exogenous. I now assume that investors have investment opportunities elsewhere in the world that yield a real return r^*. The country considered here is small in the sense that it does not affect this rate. In a perfect foresight equilibrium the level of foreign investment will ensure a capital stock that satisfies

$$r^*=G_K(K^*,L,1), \tag{18}$$

---

is taken to be anything more than five years, then the corresponding annual interest rate is near zero, suggesting that the effect in the United States is very likely to be positive.

so that

$$(19)\quad F_t = K^* + q^* - L\left[w_{t-1} - c^y(w_{t-1}, r^*)\right];$$

where

$$q^* = \frac{G_T(K^*, L, 1)}{r^*}.$$

A. *The World Interest Rate*

The effect of a permanent decline in the world interest rate is qualitatively the same as that of a permanent increase in exogenous net foreign investment. Hence, even a capital importer can experience a drop in steady-state welfare as a consequence of a lower interest rate. In contrast, this result cannot happen in Diamond's (1965) model when the steady state is efficient since this result requires that the lower interest rate raise the price of land.

If anticipated, the announcement of the lower rate benefits the contemporaneous retired generation if it appreciates land values. Hence the effect can never be Pareto worsening.[9]

When the decline in the world interest rate is unanticipated, if it causes the price of land to rise then the effect on the currently retired generation of nationals depends on how claims on land and capital are distributed between nationals and foreigners. If nationals own relatively more land, then improvement in their welfare is more likely.

When the decline in the world interest rate causes a drop in the price of land, in which case an improvement in steady-state welfare is assured, a decline in the welfare of the currently retired upon announcement is inevitable. This is true regardless of whether the change is anticipated in advance or a surprise on impact, and regardless of the distribution of claims.

B. *The Land-Labor Ratio*

Since savings behavior and foreign investment determine the capital stock, only land and labor are primitive endowments. Homogeneity of the production function implies that only the *ratio* of labor to land matters in determining factor rewards, individual welfare, and the amount of foreign investment per unit of land.

An increase in the labor force will raise or lower the total capital stock depending upon whether capital and labor are complements or substitutes in production ($G_{KL} \gtrless 0$). What happens to the capital-labor ratio depends upon the sign of $-G_{KT}$. Unless capital and land are substitutes, the capital-labor ratio falls when the labor force rises. More crowded countries have a lower capital stock per worker.

Whatever the effect on the capital stock, an increase in the labor-land ratio raises the return to land, and hence its price, and lowers the real wage ($w_L \leq 0$ and $\pi_L \geq 0$).[10] Steady-state welfare always falls, but when the increase in the labor force is first anticipated landowners experience a capital gain.[11]

Whether more crowded countries export or import capital is ambiguous. Differentiating expression (19), with w_{t-1} assuming its steady-state value, gives

$$(20)\quad \frac{dF}{dL} = \frac{dK^*}{dL} + \frac{\pi_L}{r^*} - (w - c^y) - (1 - c^y_w) w_L L.$$

---

[9] Note that factor prices and the domestic capital stock, and consequently the price of land, now assume their new steady-state values immediately upon any anticipated change in exogenous variables. From equation (19) the level of foreign investment therefore achieves its new steady-state value after one period.

[10] These results follow from the relationships $w_L = G_{LL} + G_{LK}(dK^*/dL)$; $\pi_L = G_{TL} + G_{TK}(dK^*/dL)$; $(dK^*/dL) = -(G_{KL}/G_{KK})$; the Euler conditions, and the condition that the principal minors of the matrix of second derivatives of the production function alternate in sign. See Samuelson (1947, p. 62).

[11] If there is one sector of the economy that employs only capital and labor, as in the specific-factors model, then factor rewards, the price of land and welfare are all unaffected by an increase in the labor force. This sector expands to employ the increase in the labor force at the initial capital-labor ratio.

Unless labor and capital are strongly substitutable ($G_{KL} \ll 0$), *domestic* wealth rises, both because the capital stock is likely to rise and because the value of land rises. The effect on *national* wealth is ambiguous. There are more workers saving, but since the wage is lower each saves less.[12]

C. *Relative Commodity Prices*

The model can also be used to consider the effect of a change in a relative commodity price on foreign investment. In the specific-factors model, for example, an increase in the relative price of the agricultural commodity raises the value of land but reduces the equilibrium level of the capital stock.[13] The net effect on domestic wealth in steady state is given by the sign of the expression

$$\frac{\pi}{\sigma w(L-L_M)}-\frac{r^*K^*}{wL^M},$$

where σ is the elasticity of substitution between land and labor in agriculture. If this elasticity is low or if the labor share in agriculture is relatively low then an increase in the price of the agricultural commodity has little effect on the allocation of labor. Its primary effect is to raise land values. This attracts foreign capital. At the opposite extremes, if σ is large and agriculture labor-intensive, then the major effect is to reduce the equilibrium capital stock, which acts to lower foreign investment.

IV. The Consequence of a Tax on Land with International Capital Mobility

In contrast to what is implied by a static model of tax incidence, that a tax on land is neutral in its effects on resource allocation, in a closed economy, overlapping generations context such a tax, by reducing the price of land, increases savings available for capital investment.[14] The steady-state supply of capital and wage rise and the interest rate falls. Steady-state welfare rises.

With capital mobility, a tax on land has no effect on factor prices or on the domestic stock of capital. These are governed by the world interest rate. Nevertheless, the tax does have real effects. It lowers the value of foreign investment and, as in a closed economy, it raises steady-state welfare.

A. *The Land Tax and Foreign Investment*

Consider a tax in amount τ that is collected on all domestic land each period. Foreign investment continues to imply a domestic capital stock that equates the world interest rate to the domestic marginal product of capital. Since factor supplies are unchanged and marginal productivity conditions continue to determine factor rewards, the wage and the return to land are unaffected.

---

[12] With a Cobb-Douglas technology and isoelastic marginal utility of consumption, the effect of an increase in the labor force on the level of foreign investment has the sign of the expression $(1-\beta)-\beta(1-\sigma)r^*$, where σ is the share of wage income consumed and β the labor share in production. Of two countries with the same technologies and preferences, the more crowded one will import capital if the labor share, the interest rate, and the savings propensity are small. Otherwise such a country will export capital. In the specific-factors model, since factor rewards do not change, national wealth rises in proportion to the labor force while domestic wealth rises by the increase in the labor force times the capital-labor ratio in manufacturing. Whether foreign investment rises or falls depends upon whether the capital-labor ratio in manufacturing is larger or smaller than national wealth per worker or, equivalently, if the share of land in national wealth is smaller or larger than the ratio of the labor force in agriculture to total labor.

[13] I assume that capital is an output solely of the manufacturing sector, which involves some loss of generality. A more complete analysis would introduce a price index for investment. Examination of the polar opposite case, an agricultural investment good, yielded very similar conclusions, however.

[14] The effect is similar to the "Tobin effect" of inflation on the capital stock: higher anticipated inflation channels savings out of cash balances into capital investment (James Tobin, 1965). Chamley and Wright (1987) emphasize the quantitatively trivial potential for inflation to affect the capital stock in contrast with the potential effect of a land tax.

The tax does affect the *price* of land, however. Equation (4) must be changed to

$$\frac{q_{t+1}+\pi_{t+1}-\tau}{q_t}=1+r_{t+1}. \tag{21}$$

If retirees receive a share λ of the tax revenue and workers the rest, then the equation of motion for the capital stock in a perfect foresight equilibrium is

$$\begin{aligned} K_{t+1}=L\Bigg\{ w_t - c^y & \\ & c^y\left[w_t+\tau\left(1-\frac{\lambda r^*}{1+r^*}\right)\Big/L, r^*\right]\Bigg\} \\ & +(1-\lambda)\tau-q_t+F_t, \end{aligned} \tag{22}$$

which, in steady state, implies that

$$\begin{aligned} F= K^*+\frac{\pi-\tau}{r^*}-(1-\lambda)\tau & \\ -L\left\{w-c^y\left[w+\tau\left(1-\frac{\lambda r^*}{1+r^*}\right)\Big/L, r^*\right]\right\}, & \end{aligned} \tag{23}$$

where K^* satisfies

$$G_K(K^*,L,1)=r^*.$$

The effect of an increase in τ on F is unambiguously negative regardless of λ. A land tax reduces foreign investment. Whatever the level of the tax, foreign investment falls more when more of the revenue is distributed to current workers (the lower λ).

B. *The Land Tax and Welfare*

However the tax revenue is distributed, the land tax does not affect the steady-state wage rate or rate of return on savings. But each individual receives, in addition to his wage income, redistributed tax proceeds with value $\tau[1-\lambda r^*/(1+r^*)]$. Consequently, a tax on land raises steady-state welfare regardless of how the revenue is distributed between generations.

Steady-state welfare rises as τ rises. The potential welfare gain is not, of course, unlimited. If land can remain unowned then when $\tau>\bar{\pi}$ no land is owned and no tax is collected; $\bar{\pi}$ consequently imposes an upper bound on τ.[15]

While Feldstein (1977) and Chamley and Wright (1987) demonstrate that a land tax raises *steady-state* welfare in a closed economy, they also show that such a tax does not constitute a Pareto improvement. This is likely to be the case here as well. If the tax is announced in advance of its imposition, or if the tax revenue is distributed to workers at the time the tax is announced, retirees owning land experience a capital loss for which they receive no compensation. Their welfare drops because of the tax. If, however, the tax is imposed immediately upon its announcement *and* retirees are given the tax revenue, then the potential for a Pareto-improving land tax, from a national perspective, emerges.

This outcome requires that some foreign investment take the form of claims on land. Let T_0^f denote the initial amount of foreign land owned by foreigners. A sudden land tax creates an immediate capital loss for retirees in total amount $(1-T_0^f)\tau/r^*$. If $1-T_0^f>r^*$, of course, then this magnitude exceeds tax revenue, τ. Even if retired landowners are given the tax revenue, it does not compensate them for their capital loss. They are net losers. If $1-T_0^f<r^*$, however, tax receipts exceed the capital loss to nationals, and retirees are net beneficiaries of a land tax. Since steady-state welfare rises as a consequence of the tax, and the new steady state is achieved in one period, the tax yields, from a *national* perspective, a Pareto improvement. The loss on the part of foreign

---

[15] If landowners *must* sell their land, then a negative price of land must be admitted. The upper bound on τ in this case is determined by the wealth of landowners at the period in which the tax is initially imposed.

landowners is, of course, the source of the gain to the initial retirees.

In summary, while a land tax raises steady-state welfare, it typically harms the retired generation at the time it is announced. For a national Pareto improvement to emerge requires that: (*i*) the tax is not anticipated, (*ii*) tax revenue is distributed to retirees, *and* (*iii*) the share of nationally owned domestic land is less than the world interest rate.

V. Conclusion

Capital and land serve both as factors of production and as stores of value. The primary distinction between capital and land is that the supply of the first is determined by household savings and foreign investment. Nature determines the supply of land. Models of international trade and investment have typically incorporated land and capital only in their roles as factors of production, not as competing assets in household portfolios. Once a role for land as an asset is recognized, standard results on the implications of international investment can change significantly. This paper has examined the effects of international investment when land is both a factor of production and an asset in fixed supply.

Even though results, in general, are ambiguous, the direction of the effects of changes in relevant exogenous variables can be inferred from the underlying technology. Analysis of aggregate production relationships has typically not distinguished between reproducible and nonreproducible factors of production, however, so that the empirically relevant effects are difficult to ascertain. Examination of the meager data that are available suggests that, in different countries, different answers may emerge to questions about the consequences of foreign capital for welfare and land prices, and about the implications of interest rate, terms-of-trade, and resource supply changes for welfare, the price of land, and the direction of foreign investment. Much more data and analysis are needed before any answers can be provided with reasonable confidence.

APPENDIX

Equations (8) and (9) in the text may be linearized around $\bar{K}$ and $\bar{q}$ as follows:

$$\text{(A1)} \qquad x_{t+1} = L(1-c_w^y)w_K x_t - Lc_r^y r_K x_{t+1} - y_t$$

$$\text{(A2)} \qquad y_t = \left[\frac{\pi_K}{1+r} - \frac{\bar{q}}{1+r} r_K\right] x_{t+1} + \frac{1}{1+r} y_{t+1},$$

where $x_t \equiv K_t - \bar{K}$ and $y_t \equiv q_t - \bar{q}$.

These equations constitute a second-order system of linear homogeneous difference equations. Their solution has the form

$$\text{(A3)} \qquad x_t = A_1(\rho_1)^t + A_2(\rho_2)^t$$

$$\text{(A4)} \qquad y_t = \frac{\rho_1 - \Gamma_{11}}{\Gamma_{12}} A_1(\rho_1)^t + \frac{\rho_2 - \Gamma_{11}}{\Gamma_{12}} A_2(\rho_2)^t,$$

where ρ_1 and ρ_2 are roots of the equation

$$\rho^2 - (\Gamma_{11} + \Gamma_{22})\rho + \Gamma_{11}(1+r) = 0,$$

where

$$\Gamma_{11} \equiv \frac{L(1-c_w^y)w_K}{1+Lc_r^y r_K}$$

$$\Gamma_{12} \equiv \frac{-1}{1+Lc_r^y r_K}$$

$$\Gamma_{21} \equiv -(\pi_K - \bar{q} r_K)\Gamma_{11}$$

$$\Gamma_{22} \equiv 1 + r + \Gamma_{12}\Gamma_{21}/\Gamma_{11},$$

and with ρ_1 and ρ_2 defined so that $\rho_1 \leq \rho_2$. The scalars A_1 and A_2 are determined by boundary conditions.

At any period t the capital stock K_t, and hence x_t, are predetermined. The price of land q_t, and hence y_t, are determined by asset market equilibrium each period. At period $t=0$, A_1 must therefore satisfy the condition $A_1 = x_0 - A_2$.

A necessary and sufficient condition for the existence of a unique stable nonoscillating convergence path to steady state is that $0 \leq \rho_1 < 1 \leq \rho_2$.

If $\rho_1 < 0$ the system oscillates. The conditions $\Gamma_{11} > 0$ and $\Gamma_{11} + \Gamma_{22} > 0$ preclude oscillation. These conditions in turn are guaranteed by the two conditions

$$1 \geq c_w^y, \tag{A5}$$

which states that the marginal propensity to save is nonnegative, and

$$1 + Lc_r^y r_K > 0, \tag{A6}$$

which requires that an increase in the interest rate not have a highly negative effect on savings.

If $\rho_1 \geq 0$ and $\rho_2 < 1$ then for *any* value of A_2 equations (A3) and (A4) converge to steady state while if $\rho_1 \geq 1$ no path converges. A unique convergence path therefore requires that $0 \leq \rho_1 < 1 \leq \rho_2$ and $A_2 = 0$.

If (A5) and (A6) are satisfied, then a necessary and sufficient condition for a unique stable nonoscillating convergence path is that $\Delta \geq 0$, where Δ is defined in equation (14).

To verify that $\Delta \geq 0$ is necessary and sufficient define the function

$$\Lambda(\rho) \equiv \rho^2 - (\Gamma_{11} + \Gamma_{22})\rho + \Gamma_{11}(1+r).$$

Since $\Lambda(\rho) = 0$ defines ρ_1 and ρ_2, for $\rho \in (\rho_1, \rho_2)$, $\Lambda(\rho) < 0$. If $\rho_1 < 1 \leq \rho_2$, then $\Lambda(1) \leq 0$. Since

$$\Lambda(1) = \frac{r\Delta}{-(1 + Lc_r^y r_K)},$$

along with (A6), $\Delta \geq 0$ implies and is implied by $\Lambda(1) \leq 0$.

REFERENCES

Buiter, Willem H., "Time Preference and International Lending and Borrowing in an Overlapping-Generations Model," *Journal of Political Economy*, August 1981, *89*, 769–97.

Calvo, Guillermo A., Kotlikoff, Laurence and Rodriguez, Carlos Alfredo, "The Incidence of a Tax on Pure Rent: A New (?) Reason for an Old Answer," *Journal of Political Economy*, August 1979, *87*, 869–74.

Chamley, Christophe and Wright, Brian D., "Fiscal Incidence in Overlapping Generations Models with a Fixed Asset," *Journal of Public Economics*, February 1987, *32*, 3–24.

Denison, Edward F., *Accounting for United States Economic Growth*, Washington, D. C.: The Brookings Institution, 1974.

Diamond, Peter A., "National Debt in a Neoclassical Growth Model," *American Economic Review*, December 1965, *55*, 1126–50.

Eaton, Jonathan, "A Dynamic Specific-Factors Model of International Trade," *Review of Economic Studies*, April 1987, *54*, 325–38.

Feldstein, Martin, "The Surprising Incidence of a Tax on Pure Rent: A New Answer to an Old Question," *Journal of Political Economy*, April 1977, *85*, 349–60.

Galor, Oded, "Time Preference and International Labor Migration," *Journal of Economic Theory*, February 1986, *38*, 1–20.

George, Henry, *Progress and Poverty*, New York: Doubleday Page and Co., 1904.

______, *The Science of Political Economy*, New York: Doubleday and McClure, 1898.

Goldsmith, Raymond W., *Comparative National Balance Sheets*, Chicago: University of Chicago Press, 1985.

Jones, Ronald W., "A Three-Factor Model in Theory, Trade and History," in Jagdish N. Bhagwati, ed., *Trade, Balance of Payments and Growth*, Amsterdam: North-Holland, 1971, 3–21.

Kareken, John and Wallace, Neil, "Portfolio Autarky: A Welfare Analysis," *Journal of International Economics*, February 1977, *7*, 19–43.

MacDougall, G. D. A., "The Benefits and Costs of Private Investment from Abroad: A Theoretical Approach," *Economic Record*, March 1958, *36*, 13–35.

Ricardo, David, *On the Principles of Political Economy and Taxation*, London: John Murray, 1817. Reprinted in *The Works and Correspondence of David Ricardo*, Pierro Sraffa, ed., Vol. I, New York: Cambridge University Press, 1951.

Ruffin, Roy J., "International Factor Movements," in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, Vol. I, Amsterdam: North-Holland, 1984, 237–88.

Samuelson, Paul A., *Foundations of Economic Analysis*, Cambridge: Harvard University Press, 1947.

______, "Ohlin was Right," *Swedish Journal of Economics*, December 1971, *73*, 365–84.

Tirole, Jean, "Asset Bubbles and Overlapping Generations," *Econometrica*, November 1985, *53*, 1499–1522.

Tobin, James, "Money and Economic Growth," *Econometrica*, October 1965, *33*, 671–84.

World Bank, *World Development Report* 1980, Washington, D. C.

New Estimates of Quality of Life in Urban Areas

By GLENN C. BLOMQUIST, MARK C. BERGER, AND JOHN P. HOEHN*

Implicit markets capture compensation for intraurban and interregional differences in amenities and yield differences in housing prices and wages. These pecuniary differences become preference-based weights in a quality of life index. Hedonic equations are estimated using micro data from the 1980 Census and assembled county-based amenity data on climatic, environmental, and urban conditions. Ranking of 253 urban counties reveals substantial variation within and among urban areas.

Quality of life comparisons among areas attract the attention of residents, workers, business managers, and policymakers. Evidence of the influence of amenities, one dimension of quality, on housing, job, and migration decisions is found by Ronald Krumm (1980), Philip Graves (1983), and V. Kerry Smith (1983) among others. People are interested in comparing the bundle of amenities available at one location to bundles elsewhere.

The basic problem in constructing a quality of life index is developing a method for weighting the different amenities. Early work ranking cities by quality of life, for example, Ben-Cheih Liu (1976), produced overall quality of life indexes by weighting amenities in an atheoretic manner. Advances were made by Sherwin Rosen (1979), who suggests urban locations are best viewed as tied bundles of wages, rents, and amenities. Extending Rosen's work, Jennifer Roback (1982) constructed a model to show interregional amenity differences are bid into interregional differences in both wages and land rents. These differences yield implicit amenity prices which are appropriate weights for a quality of life index. The quality of life index we develop incorporates the wage and rent effects discussed by Rosen and Roback.

In contrast to previous research, our model allows for amenity variation both within and across urban areas. Agglomeration effects due to the productivity effects of city size provide a key linkage between firms of a given urban area. Our data extend the reach of empirical analysis; more cities are included, amenities are measured at the county level, housing rent data are for individual households rather than urban aggregates, and the data are more recent. Hedonic wage and rent equations are estimated using 1980 U.S. Bureau of the Census micro data matched with amenity data on climatic, environmental, and urban conditions. Implicit amenity prices are estimated from the hedonic results. These implicit prices are the amenity weightings in our quality of life index. The index is computed for 253 urban counties within 185 metropolitan areas of the United States.

I. Framework

A. *Households, Firms, and Urban Structure*

Central to our framework is the idea that different urban locations offer different sets of wages, rents, and amenities. Households attempt to maximize well-being and firms attempt to minimize costs by their location decisions. We assume a fixed number of urban areas in which individuals and firms may locate. Before location decisions are made, households and firms are freely mobile.[1] Each urban area is a collection of counties. For simplicity, we analyze the case of an arbitrary urban area composed of two counties. Each county is composed of a fixed amount of land and offers a different set of amenities that resident households and firms may enjoy. Counties in an urban area are linked together by agglomeration effects; the population of the entire urban area affects the production costs of firms, regardless of the county in which they are located.

---

*Blomquist and Berger, Department of Economics, University of Kentucky, Lexington, KY 40506-0034, and Hoehn, Department of Agricultural Economics, Michigan State University, East Lansing, MI 48824-1039. Helpful comments on earlier drafts were made by Timothy Bartik, Imdad Hussain, Jan Owen Jansson, Ronald Krumm, Alan Randall, V. Kerry Smith, George Tolley, three anonymous referees, and participants in seminars at the University of Chicago, University of Kentucky, University of Newcastle Upon Tyne, University of Stockholm, and Vanderbilt University. We gratefully acknowledge initial support from USEPA under Cooperative Agreement 811056-01-0. This research was supported also in part with funds from the National Science Foundation (grant no. RII-8610671) and the Commonwealth of Kentucky through the Kentucky EPSCoR Program. The views expressed are the authors' alone. Research assistance was provided by Michael Alexander.

Households gain utility through use of a composite commodity, local residential land, and local amenities. A household gains access to the amenities of the kth county through the purchase of residential land q_k, where $k \in \{1,2\}$. Both land and the consumption commodity are purchased out of labor earnings. Each household is endowed with one unit of labor. A household in county k sells its labor time to local firms and earns a wage w_k. Earnings comprise all of income and labor is homogeneous. Labor transportation costs within a given county are assumed to be negligible.[2]

The level of well-being attained by a household in county k is

$$v^k = v^k(w_k, r_k; a_k), \tag{1}$$

where $v^k(\cdot)$ is the indirect utility function, r_k is the rental price of land in the kth urban area county, a_k is an index of local amenities, and the fixed, unit price of the composite commodity is suppressed. Utility in the kth county increases with an increase in wages in the kth county ($v^k_{w_k} > 0$) and decreases with an increase in the rental price of land within the county ($v^k_{r_k} < 0$). An increase in a county's amenities increases utility in that county if a_k is a consumption amenity ($v^k_{a_k} > 0$) and decreases utility if a_k is a disamenity in consumption ($v^k_{a_k} < 0$). A household residing in county k demands residential land $q_k = v^k_r / v^k_w$. Letting the amount of land within the kth county be fixed and equal to Q_k, the population of county k is $N^k = Q_k / q_k$. Given equation (1), this implies that a county's population is a function of its own wage, rent, and amenity levels.

Firms combine local labor and capital to produce the composite commodity. The prices of capital and the composite commodity are fixed by international markets. Prices and wages are normalized on the price of the composite good, and the price of the composite good is set to unity. Production technology is constant returns to scale in labor and capital. Local amenities enter production as shift parameters.

For a firm located in the kth county, unit production costs are

$$c^k = c^k(w_k; a_k, N), \tag{2}$$

---

[1] The intraurban, interregional model is based on the assumption that the housing market is sufficiently national in scope that compensating differentials surface in the housing market as well as in the labor market. For evidence that is suggestive of the appropriateness of the national housing market assumption see Peter Linneman (1980).

[2] For simplification we assume there is no cross-county commuting, work hours are exogenous, and labor is homogeneous. Of course, commuting does exist, but most people do live and work in the same county. As calculated from the migration—place of work—travel time subsample of the 1980 Census A sample, among the people who live in one of the 340 urban counties identified, 81 percent live and work in the same county.

While labor supply is exogenous in our model, there may be some labor market response by households to amenity-level differences. We expect households who choose areas with low wages and high amenity levels to reduce their hours worked in response to the low relative price of leisure. This partial effect in the labor market magnifies the reduction in wages necessary to compensate for a difference in amenities. Even though we do not model labor supply or bring it directly into the estimation, we do illustrate how our quality of life measure varies with differences in hours worked by the household.

Another complication which has been considered in detail elsewhere is heterogenous labor. Roback (1988) shows that in an interregional context diversity of workers further dilutes the simple presumption of a negative relationship between amenities and wages and a positive relationship between amenities and rents.

where $c^k(\cdot)$ is a firm's unit cost function and the price of capital is left implicit. N is the population size of the entire urban area. N reflects agglomeration or congestion effects (George Tolley, 1974). As city size varies, unit costs of a firm may fall, remain constant, or increase ($c_N \gtreqless 0$). In terms of amenities, unit costs decrease with a change in a_k, if a_k is an amenity in production ($c^k_{a_k} < 0$), and increase with a change in a_k, if a_k is a production disamenity ($c^k_{a_k} > 0$). By Shephard's lemma a firm's demand for labor is $c^k_{w_k} > 0$.

As individuals and firms locate across different urban areas and counties, wages and rents adjust to clear their respective markets. A spatial equilibrium implies that households cannot improve utility and firms cannot reduce their costs by relocating. An intraurban equilibrium occurs when firms within an urban area produce at common production costs and all urban area households attain an equal level of utility. An interurban equilibrium implies that unit production costs are equal to the unit product price and households across all urban areas attain a common level of utility, u^0.

For an arbitrary urban area, the set of wages, land rents, and city size that sustains an intraurban and interurban equilibrium satisfies the system of equations

$$\text{(S1)} \qquad 1 = c^k(w_k; a_k, N)$$

$$u^0 = v^k(w_k, r_k; a_k)$$

$$N = \sum_{k=1}^{2} N^k,$$

where $k \in \{1,2\}$. The last equation links equilibrium wages and rents within an urban area through the effect of city size on firms' productivity. Using the set of wages and rents that satisfy (S1), one can solve for (a) the implicit price of the amenity a_k, and (b) the comparative static effects of a change in a_k on the equilibrium set of wages, land rents, and city size.

B. *Wages, Rents, and Implicit Amenity Prices*

Equilibrium wage and land rent differentials are used to compute the implicit prices of amenities. These implicit prices, f_k, are obtained by taking the total differential of the second equation in (S1) and rearranging to find $f_k = v^k_{a_k}/v^k_{w_k}$. Consequently, the implicit price of a_k is

$$\text{(3)} \qquad f_k = q_k(dr_k/da_k) - dw_k/da_k,$$

where dr_k/da_k is the equilibrium rent differential and dw_k/da_k is the equilibrium wage differential. We have Roback's (1982) result—that the implicit price is the sum of the land expenditure differential and the negative of the wage differential.

Comparative static analysis is used to solve for the anticipated sign of the equilibrium wage, rent, and city-size differentials, given specific assumptions regarding a_k. Unlike Roback we allow amenities to vary within an urban area; we permit $a_1 \neq a_2$ and $da_1 \neq da_2$. Comparative static results are obtained by taking the total differential of the system (S1) and then solving for the appropriate differentials. From the comparative static analysis, the wage and rent differentials are

$$\text{(4)} \quad dw_k/da_k = \Big(- v^k_{a_k} c^k_N c^j_{w_j} v^j_{r_j} N^k_{r_k} - c^k_{a_k} v^k_{r_k} v^j_{r_j} \beta_j + N^k_{a_k} v^k_{r_k} v^j_{r_j} c^k_N c^j_{w_j} \Big) / D$$

$$\text{(5)} \quad dr_k/da_k = \Big[- v^k_{a_k} v^j_{r_j} \big(c^k_{w_k} \beta_j - c^j_{w_j} c^k_N N^k_{w_k} \big) + c^k_{a_k} v^k_{w_k} v^k_{r_k} \beta_j - N^k_{a_k} v^k_{w_k} v^j_{r_j} c^k_N c^j_{w_j} \Big] / D,$$

where $k \neq j \in \{1,2\}$, $\beta_j = c^j_N[(v^j_{w_j}/v^j_{r_j})N^j_{r_j} - N^j_{w_j}] - c^j_{w_j}$, and D is the determinant of the differential system. $N^k_{r_k}$ is the partial effect of a change in rent in county k on population in county k. $N^k_{w_k}$, $N^j_{r_j}$, and $N^j_{w_j}$ are similarly defined for the wage effect and the effects in county j. $N^k_{a_k}$ is the partial effect of a change in amenities in county k on that county's population. Three more differentials exist and are shown in the Appendix. Two,

dw_k/da_j and dr_k/da_j, show the cross-county effects that occur between two counties within a given urban area. These arise due to agglomeration effects—a change in amenities in county j results in a change in city size that in turn affects the cost of firms within county k. The shift in the costs induces a change in wages and rents. Interestingly, the cross-county rent differential is proportional to the cross-county wage differential. The last differential, dN/da_k, indicates the effect of city size due to a change in the kth county's amenities.

The signs of the five wage, rent, and city-size differentials depend on alternative assumptions regarding: 1) the effect of amenities on households, 2) the impact of amenities on production costs, and 3) the effects of agglomeration. In general, it is not true that the wage differential is negative and the land rent differential is positive even if a_k is a consumption amenity. For instance, in the case in which the amenity increases household utility and reduces production costs, the own-county wage and rent differentials are both positive except where city-size effects increase firm costs. In the case in which city size increases firm costs and a_k is an amenity to households, but neutral with regard to unit costs, own-county rents and city size increase unambiguously with a_k. The sign of the agglomeration effect determines the sign of the wage differential. If firm costs fall as city size increases, the wage differential is positive. If size diseconomies prevail, the wage differential is negative. In the case in which a_k is an amenity to households and detrimental to firms, most of the comparative static effects are ambiguous. The comparative static results underscore the importance of accounting for both the wage and rent differentials in computing quality of life weights.[3]

Two problems arise in empirically implementing our model. First, the amenity price given in equation (3) is based in part upon residential land rents. However, residential land is typically bundled into a package of housing characteristics and land rents are difficult to observe. We can extend our model by introducing housing instead of land in the utility function and adding a production function for housing. As can be shown, the amenity price equation becomes

$$(6) \qquad f_k = h_k(dp_k/da_k) - dw_k/da_k,$$

where h_k is the quantity of housing purchased by a household in county k and p_k is the price of housing in county k.

Second, an estimate of equation (6) appears to require an unbiased estimate of own-county wage and housing-price differentials. In general, unbiased estimates of these wage and rent differentials would require controlling for the levels of amenities prevailing in all the other counties within a particular urban area. For several large urban areas in the United States this approach would require equations with an unwieldy number of independent variables. The omission of the cross-county amenities leads to biased own-county wage and rent differentials if amenities are correlated across counties in the same city. It can be shown, however, that in the amenity price equation (6), the terms causing the bias for each differential cancel.[4] The estimate of the full amenity price is unbiased and useful as a weight in our quality of life index.

---

[3] In John Hoehn, Mark Berger, and Glenn Blomquist (1987) we show that in an interregional context with urban structure, full amenity values are a weighted average of compensation changes in housing and labor markets. In that paper, which focuses on differences between partial and full prices of amenities, monotonically increasing commuting costs are the sole source of variation in housing prices within each city. In our current model we impose less urban structure and allow for changes in a vector of amenities between counties. The result that full amenity values are a combination of compensatory changes in both markets also follows from our curent model. In the current model we are able to consider (noncommuting) variation in amenity changes within a city as well as across cities in order to facilitate comparisons of quality of life.

[4] A system of equations similar to (S1) except substituting housing for land, is used to derive the own-county wage and housing-price differentials and the cross-county wage and housing-price differentials. The derivations (similar to those in the Appendix) show that the cross-county wage and housing-price differentials are strictly proportional where the reciprocal of housing consumption is the factor of proportionality. If the wage and housing equations are estimated without

II. Data

The housing expenditure and wage equations used to measure amenity values are estimated with the 1 in 1000 A Public-Use Sample of the 1980 Census and other assembled data. The non-Census data are obtained from a variety of sources and merged to the Census data by county, Standard metropolitan Statistical Area (SMSA), or industry. Together, the Census data and the merged aggregate data consist of observations on 34,414 housing units and 46,004 individuals from 253 counties across the nation.[5]

The housing sample includes all housing units on 10 or fewer acres for which value of the unit or contract rent is reported. In the wage sample are all individuals aged 16 and over who reported their earnings, hours and weeks, had nonzero wage and salary earnings, and had positive total earnings. Monthly housing expenditures is the dependent variable in the housing equation. For renters, monthly housing expenditures is defined as gross rent including utilities. For owners, reported house value is converted to monthly imputed rent using a 7.85 percent discount rate obtained from a user cost study by Richard Peiser and Lawrence Smith (1985). Monthly expenditures for utilities are added to obtain gross imputed rent for owners. The dependent variable in the wage equation is average hourly earnings. The variable is obtained by dividing annual earnings by the product of weeks worked during the year and usual hours worked per week.

The housing-hedonic regression includes Census measures of structural characteristics and also central city status. The wage-hedonic regression has Census-based variable controlling for personal characteristics, occupational group, and central city status. Included in the wage equation is industry unionization from Edward Kokkelenberg and Donna Sockell (1985). The remaining variables common to both the housing- and wage-hedonic regressions come from data merged with the 1980 Census. These variables represent fifteen of the sixteen components of our quality of life index (QOLI). There are six variables measuring climatic conditions taken from *Comparative Climatic Data* prepared by the National Climatic Data Center (1983). A violent crime variable is included based on data reported in the *U.S. FBI Uniform Crime Reports for the United States* (1980). Coast is a variable that equals one if the county of residence touches an ocean or one of the Great Lakes. The teacher-pupil ratio is based on school district and county data on enrollment and salaries found in Volumes 3 and 4 of the 1982 *Census of Governments*. This school-input ratio is similar in construction to the ratio used by Steven Cobb (1984).

The estimated wage and housing-expenditure equations also include six environmental variables based on data from four differ-

---

amenities from other counties, then the estimated own-effect differentials (γ_k^w and γ_k^p) each contain additional terms

$$\gamma_k^w = dw_k/da_k + \sum_{j \neq k} \delta_{kj}(dw_k/da_j),$$

$$\gamma_k^p = dp_k/da_k + \sum_{j \neq k} \delta_{kj}(dp_k/da_j),$$

where the a_j's are the levels of amenity a in other counties in the city and δ_{kj}'s are the coefficients obtained from a regression of a_k on the a_j's (J. Johnston, 1984, p. 260). However, use of γ_k^w and γ_k^p results in an estimate of f_k as

$$\begin{aligned} h^k\gamma_k^p - \gamma_k^w \\ &= h^k\left[dp_k/da_k + \sum_{j \neq k} \delta_{kj}(dp_k/da_j)\right] - dw_k/da_k \\ &\quad - \sum_{j \neq k} \delta_{kj}(dw_k/da_j) \\ &= h^k(dp_k/da_k) - dw_k/da_k = f_k \end{aligned}$$

since from the Appendix $dp_k/da_j = (1/h_k)dw_k/da_j$. While the above refers to the case of a single amenity, the same result holds for multiple amenities where there is correlation across amenities. Thus, our empirical work provides unbiased estimates of the full amenity prices in addition to being conceptually superior to studies based on a single market.

[5] Although we have a national sample, we do not include all counties in the United States for two reasons. First, the Census A sample identifies only 340 counties, all of which have a population exceeding 100,000. Second, the number of counties is reduced further because we do not have a full set of amenity data for 87 of 340 urban counties.

ent sources. The ambient concentration of total suspended particulate (TSP) for each county is based on the U.S. Environmental Protection Agency (USEPA) SAROAD data. Visibility data are obtained from reports by John Trijonis and Dawn Shapland (1979) and daily weather observations are supplied by the National Climatic Center. The number of Superfund sites in the individual's county is based on information published in the Council on the Environmental Quality report, *Environmental Quality 1982*. The last three variables are measures of activity in the individual's county of residence based on the Resource Conservation and Recovery Act (RCRA) Application for Hazardous Waste Permit Tape obtained from the USEPA. The first two are counts of the number of treatment, storage, and disposal (TSD) facilities for hazardous wastes, and the number of National Pollution Discharge Elimination Systems (NPDES) water-pollution discharges in the county of residence. The last variable is the total licensed waste for landfills in the county.

During the course of the project we considered including other amenity measures in our overall bundle. Some of them, such as number of sports teams, were too highly correlated with an endogenous variable (city size) in our model. Others, such as overall tourism expenditures, had a price component as well as a quantity component, which made them unattractive. Finally, several measures could not be collected for a sufficient number of counties to estimate a national model which uses those measures.

III. Hedonic Regressions, Full Implicit Prices, and Amenity Components

Housing and wage hedonic equations are estimated for households and workers. A Box-Cox search was done over functional forms of

$$(7)\qquad \frac{Y^{\lambda}-1}{\lambda}=b_0+\sum_{i=1}^{n} b_i\frac{X_i^{\gamma}-1}{\gamma}+\varepsilon,$$

where Y is either rent or wage, the X_i are the independent variables, λ was varied from $+1.2$ to -0.4, and γ was either 1 (linear) or 0 (logarithm). The best functional form for the housing hedonic was $\lambda=0.2$ and $\gamma=1$, and the best form for the wage hedonic was $\lambda=0.1$ and $\gamma=1$. These functional forms yielded the highest values of the log-likelihood functions. The log-likelihood values for the best housing and wage equations are significantly different from the double-linear, double-log, and semilog forms at the .01 level.

Table 1 shows linearized (unboxed) coefficient estimates and standard errors for the sixteen amenities and disamenities included in the housing-expenditure and wage-hedonic regressions.[6] Amenities clearly affect housing expenditures and wages. All of the amenity parameter estimates shown in column 2 are at least two times larger than their standard errors in the housing-expenditure equation. Ten out of the sixteen wage equation estimates shown in column 3 exceed two times their standard error. Compared to regressions without amenities (not shown) the addition of the amenities results in an increase in R^2 from .587 to .662 (F value $=$ 480.0) in the housing expenditure equation and from .308 to .314 (F value $=$ 25.0) for the wage equation.

The amenity coefficients are adjusted to reflect mean annual household housing expenditures and labor earnings and then are combined using equation (6) to generate full implicit prices for the amenities. The fourth column of Table 1 gives the annual full implicit price and standard error for each amenity for the average household. A positive full implicit price indicates a marginal net amenity and a negative price indicates a marginal net disamenity. Only visibility has an unexpected sign, but its standard error is

---

[6] The full set of housing-expenditure and wage-hedonic regression results are available upon request. The regression coefficients are linearized in the following manner: $b_i' = b_i(y^{-1-\lambda})$, where b_i' is the linearized coefficient, b_i is the estimated coefficient from the best functional form, $\bar{y}$ is the sample mean hourly earnings ($8.04) or monthly housing expenditures ($462.93), and λ is .2 in the housing equation and .1 in the wage equation. See Linneman (1980) for details on transforming Box-Cox coefficients.

TABLE 1—LINEARIZED PARAMETER ESTIMATES, FULL IMPLICIT PRICES, AND QUALITY OF LIFE INDEX COMPONENTS

Amenity Variable (Wage Sample Mean and Unit of Measurement)	Monthly Housing Expenditure Equation[a]	Hourly Wage Equation[b]	Full Implicit Price[c]	QOLI Component Mean	QOLI Component Minimum	QOLI Component Maximum
Precipitation	−1.047	−.0144	$23.50	$808	$88	$1574
(32.0 inches per year)	(.149)	(.004)	(9.71)	(291)		
Humidity	−2.127	.0065	−43.42	−2987	−3397	−1368
(68.3 percent)	(.251)	(.006)	(16.29)	(282)		
Heating Degree Days	−.0136	−.0001	−.08	−370	−780	−16
(4326 per year)	(.001)	(.00003)	(.06)	(167)		
Cooling Degree Days	−.0760	−.0002	−.36	−448	−1461	−20
(1162 per year)	(.002)	(.0001)	(.13)	(324)		
Wind Speed	11.88	.0961	−97.51	−881	−1209	−595
(8.89 miles per hour)	(.867)	(.022)	(55.54)	(139)		
Sunshine	2.135	−.0091	48.52	2929	2183	4172
(61.1 percent of possible)	(.235)	(.006)	(15.43)	(370)		
Coast	32.51	−.0310	467.72	105	0	468
(.330, =1 if county on coast)	(2.47)	(.063)	(161.09)	(196)		
Violent Crime	.0434	.0006	−1.03	−602	−2202	65
(647 per 100,000 population per year)	(.003)	(.0001)	(.19)	(326)		
Teacher-Pupil Ratio	635.3	−5.451	21,250.00	1774	742	4483
(.0799 teachers per pupil)	(71.6)	(1.85)	(4698.33)	(369)		
Visibility	−.8302	−.0026	−3.41	−62	−273	−27
(15.8 miles)	(.110)	(.003)	(7.03)	(55)		
Total Suspended Particulates	−.5344	−.0024	−.36	−25	−60	−13
(73.2 micrograms per cubic meter)	(.058)	(.001)	(3.77)	(7)		
NPDES Effluent Dischargers	−7.458	−.0051	−76.68	−74	−844	0
(1.51 county)	(.461)	(.012)	(30.56)	(136)		
Landfill Waste	.0095	.0001	−.11	−20	−1410	0
(477 hundred million metric tons per county)	(.001)	(.00002)	(.05)	(128)		
Superfund Sites	13.42	.1069	−106.07	−66	−955	0
(.883 per county)	(.693)	(.017)	(43.70)	(129)		
Treatment, Storage, and Disposal Sites	.2184	.0013	−.58	−9	−133	0
(46.4 per county)	(.024)	(.001)	(1.56)	(14)		
Central City	40.75	−.4537	645.02	113	0	645
(.290, =1 if residence in central city)	(2.54)	(.065)	(165.09)	(188)		
Quality of Life Index	–	–	–	186	−1857	3289
(1980 dollars per year)	–	–	–	(667)		
R^2	.6624	.3138				
n	34,414	46,004				
Log-Likelihood Value	−219,013	−124,403				

Note: Standard errors or standard deviations evaluated at the mean wage or rent and shown in parentheses.

[a] The dependent variable is actual or imputed monthly housing expenditures. Control variables which are included in the housing hedonic regression, but which are not reported include: units at address, age of structure, stories, rooms, bedrooms, bathrooms, condominium status, central air, sewer, lot size exceeds 1 acre, renter status, and renter interaction terms for each of these variables.

[b] The dependent variable is annual earnings divided by the product of annual weeks worked and usual hours per week. Control variables which are included in the wage-hedonic regression, but which are not reported include: experience, (age-schooling-6), experience squared, gender interaction with experience and experience squared, race, gender, gender interaction with race, marital status, gender interaction with marital status, gender interaction with children under 18, schooling, disabled, school enrollment status, dummies for 5 of 6 broad occupation groups and percent of industry covered by unions.

[c] As shown in equation (6) the full implicit price is the sum of the annual housing expenditure and wage differentials. To obtain an annual household full implicit price, the housing coefficients are multiplied by 12 (months per year) and the wage coefficients are multiplied by (1.54)(37.85)(42.79), the product of the sample means of workers per household, hours per week, and weeks per year.

more than twice as large as its full implicit price. Part of the difficulty is that visibility is highly correlated with humidity, total suspended particulates, and precipitation. The full implicit prices of heating degree days, total suspended particulates and treatment, storage and disposal sites are also small relative to their standard errors. The remaining twelve full implicit prices, however, are all twice as large as their standard errors.[7]

The full implicit prices in Table 1 are used to construct a quality of life index across urban counties in the following manner:

$$\text{QOLI}_k = \sum_{i=1}^{16} f_i a_{ki} \qquad k = 1, \ldots, 253, \tag{8}$$

where f_i is the full implicit price of amenity i, and a_{ki} is the quantity of amenity i observed in county k. The index is measured in dollars and is useful for making quality of life comparisons across areas. The difference in the value of the index between two counties gives the premium that the average household pays implicitly through the labor and housing markets to live in the more amenable county.[8]

Differences in the values of the QOLI are due to differences in the components, the $f_i a_{ki}$. To provide some idea of the magnitudes of these components and their variation across counties, the last three columns of Table 1 show the mean value of each amenity component across the counties in our sample, and also the standard deviation, minimum and maximum of the components. For example, individuals living in the county with the most crime are compensated $2267 implicitly in the labor and housing markets over those living in the county with the least crime, and $1600 over those in the county having the average rate of violent crime. For the entire bundle of amenities the difference in compensation between the most and least desirable county is $5146. The mean value of the QOLI is $186 while the standard deviation across counties is $667.

---

[7] Our intraurban, interregional amenity model does imply wages affect rents and rents affect wages. We estimate the wage and housing-price hedonics in reduced forms. Wages are not included in the housing-price hedonic so that dp_k/da_k implicitly reflects variation in w_k. Rents are not included in the wage hedonic so that dw_k/da_k implicitly reflects variation in p_k. These are appropriate specifications because according to our model the total, not partial, differentials belong in the amenity price, f_k, given by equation (6).

Of course, it is possible that the errors in the wage and housing-price hedonics are correlated due to unobservables which affect both wages and housing prices. We examine the extent of this correlation with a subsample of 21,379 householders, for whom we have a one-for-one correspondence between wages and housing prices. In this subsample, the correlation between the wage and housing-price residuals is .128, which is small but significant. However, seemingly unrelated regression estimates for the amenities differ inside the third decimal place from those obtained using ordinary least squares in only 3 of 32 cases. Further, the QOLI ranking obtained using the seemingly unrelated regression estimates is highly correlated with the one obtained using OLS estimates; the rank correlation is .997.

[8] The premium equals the value households place on differences in amenity bundles and is accurate for small variations in the typical bundle. Since the QOLI is based on marginal implicit prices, the estimated premia for large differences between amenity bundles are approximations. By comparing the bid function and hedonic function for a single amenity we can demonstrate that: 1) the value of the amenity is overstated for all amenity levels except the level chosen by the representative person; 2) QOLI is overestimated for high- and low-amenity counties; but 3) the county ranking by QOLI is unaffected. The ordering is not changed by overestimates for below and above average counties. For more than one amenity, the QOLI values for counties with a combination of extremely high and extremely low amenities are overestimated and these counties may be overranked.

We do find some evidence of overranking. For counties with either one amenity level more than two standard deviations below average and another amenity level more than two standard deviations above average or two amenity levels more than two standard deviations below average and two amenity levels more than two standard deviations above average, the average QOLI is $517. For counties with no amenity levels more than two standard deviations from the mean, the average QOLI is $129. Thus the counties with a combination of extremely high- and low-amenity endowments do have above average QOLI values and could be overranked. Fortunately, few counties have extreme combinations. Only 12 (5 percent) counties have a one-one extreme combination and only four (2 percent) counties have a two-two extreme combination of amenity levels. Of the 253 counties, 152 (60 percent) have no extreme values at all.

IV. Ranking Urban Areas by Quality of Life

Variation in amenity endowments across urban areas produces variation in the values of QOLI. Table 2 reports the complete ranking by quality of life for all counties for which we have complete data, 253 urban counties in 185 SMSA's. Given for each county are the rank, SMSA, and state in which the county is located and the QOLI premium. The top-ranked county is Pueblo, Colorado, with a $3,289 premium. The county with the (approximate) average QOLI is Woodbury in Sioux City, Iowa, with a $189 premium. The bottom-ranked county is St. Louis City in St. Louis, Missouri, with a $-1,857 premium.[9]

An advantage of this county-based analysis is that investigation of variation in quality of life within urban areas is possible. For the 253 counties in our sample we have observations from more than one county for 38 SMSA's. Table 3 shows the 10 SMSA's with the widest range of QOLI values. The difference in QOLI values between counties in the Milwaukee SMSA is $2,350. Milwaukee County has a QOLI of $558 which ranks 68th and Waukesha County has a QOLI of $-1,792 and is ranked 252nd. In the St. Louis SMSA, St. Louis City has a QOLI of $-1,857 which is the lowest of the 253 counties, and St. Clair, Illinois, has a QOLI of $-251 which is 195th. Other SMSA's which display intraurban differences of $900 or more are Denver-Boulder, Norfolk-Virginia Beach-Portsmouth, Minneapolis-St. Paul, Detroit, Richmond, Dayton, and Chicago. As Table 3 illustrates, substantial variation exists in quality of life within as well as across urban areas.

The QOLI values presented in Tables 2 and 3 are calculated using the characteristics of the average household and all 16 amenities included in the empirical analysis. For households having other than the mean characteristics, for example, working more or less hours, the standard QOLI may be inaccurate. Some households may value only a subset of the amenity bundle which means a more limited QOLI would be relevant for them. Also, the QOLI calculated in the paper is based on an equilibrium model, and to the extent disequilibrium influences are present a county's ranking may change. Table 4 addresses these issues by presenting alternative QOLI rankings for households with high and low hours worked and for subsets of amenities, along with giving some indication of how fast county population is growing. The alternative rankings and county population growth are shown for the top 15, middle 15, and bottom 15 counties as ranked by the standard QOLI.

The QOLI rankings for households with one-half and twice as many hours worked as the average show the sensitivity of the rankings to differences in exogenous household labor supply. Households with high and low numbers of hours worked will place different values on the amenities due to variation in the contribution of the hourly wage effect (dw/da) to the annual full implicity price. The rankings do change somewhat as hours of work change. But in general top counties stay near the top under different assumptions about hours worked and bottom counties stay near the bottom. Indeed, the rank correlation coefficients between the standard ranking and the low and high hours worked rankings are .63 and .83.

Also presented in Table 4 are rankings for three subsets of amenities: climate, environmental quality, and urban conditions. The urban conditions subset includes central city status, the violent crime rate, and teacher-pupil ratio. The rankings based on these subsets are related to the standard ranking in various degrees: the rank correlation between the climate subset and the standard

---

[9] Given our emphasis on intraurban variation in quality of life and our county-based data set, it is difficult to compare our results with SMSA rankings reported in previous studies. However, we can obtain an SMSA QOLI value by calculating population weighted average of QOLI values for counties within the SMSA. We have done this for the set of cities which are common to our study and those of Roback (1982) and Liu (1976) as reported by Roback (1982, p. 1275). The rank correlation for 19 common cities between our SMSA QOLI ranking and Roback's ranking is +0.40. The rank correlation for 18 common cities as reported by Roback between our SMSA QOLI ranking and Liu's ranking for environmental factors is +0.42. A comparison with the Rosen (1979) results is not made because he does not provide an overall ranking.

TABLE 2—QUALITY OF LIFE RANKING FOR 253 URBAN COUNTIES, 1980

Rank	County	Standard Metropolitan Statistical Area	State	QOLI (1980 dollars)
1	Pueblo	Pueblo	CO	$3288.72
2	Norfolk City	Norfolk-VA Beach-Portsmouth	VA	2105.77
3	Arapahoe	Denver-Boulder	CO	2097.07
4	Bibb	Macon	GA	1599.57
5	Washoe	Reno	NV	1575.37
		(Mean QOLI + 2 S.D. = $1520)		
6	Broome	Binghamton	NY	1485.63
7	Hampton City	Newport News-Hampton	VA	1444.63
8	Sarasota	Sarasota	FL	1430.84
9	Palm Beach	West Palm Beach-Boca Raton	FL	1422.54
10	Pima	Tucson	AZ	1341.86
11	Broward	Fort Lauderdale-Hollywood	FL	1326.91
12	Boulder	Denver-Boulder	CO	1319.47
13	Larimer	Fort Collins	CO	1297.84
14	Denver	Denver-Boulder	CO	1295.25
15	Charleston	Charleston-North Charleston	SC	1280.21
16	Monterey	Salinas-Seaside-Monterey	CA	1213.97
17	Roanoke City	Roanoke	VA	1129.65
18	Lackawanna	Northeast Pennsylvania	PA	1127.43
19	Leon	Tallahasee	FL	1066.51
20	Richmond City	Richmond	VA	1059.96
21	Fayette	Lexington-Fayette	KY	1055.50
22	Santa Barbara	Santa Barbara-Santa Maria-Lompoc	CA	1025.76
23	Ventura	Oxnard-Simi Valley-Ventura	CA	1022.83
24	Durham	Raleigh-Durham	NC	1014.01
25	New Hanover	Wilmington	NC	1000.92
26	Wake	Raleigh-Durham	NC	990.98
27	San Diego	San Diego	CA	980.93
28	VA. Beach City	Norfolk-VA Beach-Fortsmouth	VA	967.70
29	Lancaster	Lancaster	PA	965.38
30	Manatee	Brandenton	FL	958.13
31	Weld	Greeley	CO	957.23
32	El Paso	El Paso	TX	923.02
33	Racine	Racine	WI	912.83
34	Guilford	Greensboro-Winston-Salem-High Pt.	NC	908.74
35	Lane	Eugene-Springfield	OR	884.00
36	Maricopa	Phoenix	AZ	870.69
		(Mean QOLI + 1 S.D. = $853)		
37	Lycoming	Williamsport	PA	844.97
38	Brevard	Melbourne-Titusville-Cocoa	FL	819.93
39	Fresno	Fresno	CA	808.87
40	Minnehaha	Sioux Falls	SD	803.73
41	Orange	Anaheim-Santa Ana-Garden Grove	CA	803.49
42	Lee	Fort Myers-Cape Coral	FL	803.36
43	Luzerne	Northeast Pennsylvania	PA	792.13
44	Berkshire	Pittsfield	MA	792.13
45	Forsyth	Greensboro-Winston Salem-High Pt.	NC	792.04
46	Dane	Madison	WI	791.58
47	Cumberland	Vineland-Millville-Bridgeton	NJ	787.31
48	Dauphin	Harrisburg	PA	763.67
49	Suffolk	Nassau-Suffolk	NY	757.40
50	Knox	Knoxville	TN	735.50
51	Washington	Fayetteville-Springdale	AR	733.41
52	San Joaquin	Stockton	CA	728.69
53	Adams	Denver-Boulder	CO	726.73
54	Blair	Altoona	PA	711.21
55	Hudson	Jersey City	NJ	709.16
56	Jackson	Pascagoula-Moss Point	MS	699.85
57	Spartanburg	Greenville-Spartanburg	SC	687.58

TABLE 2—CONTINUED

Rank	County	Standard Metropolitan Statistical Area	State	QOLI (1980 dollars)
58	Los Angeles	Los Angeles-Long Beach	CA	667.64
59	Greenville	Greenville-Spartanburg	SC	648.88
60	Nassau	Nassau-Suffolk	NY	620.16
61	Pinellas	Tampa-St. Petersburg	FL	610.17
62	Jefferson	Denver-Boulder	CO	608.77
63	Richmond	Augusta	GA	598.73
64	Richland	Columbia	SC	586.06
65	Aiken	Augusta	SC	585.33
66	Mobile	Mobile	AL	559.55
67	Stearns	St. Cloud	MN	558.70
68	Milwaukee	Milwaukee	WI	557.84
69	Clark	Las Vegas	NV	540.27
70	Passaic	Paterson-Clifton-Passaic	NJ	536.74
71	Montgomery	Philadelphia	PA	533.91
72	Ada	Boise City	ID	533.39
73	Vanderburge	Evansville	IN	530.66
74	Monroe	Rochester	NY	503.38
75	Chatham	Savannah	GA	498.30
76	Gaston	Charlotte-Gastonia	NC	474.70
77	Polk	Des Moines	IA	470.45
78	Albany	Albany-Schenectady-Troy	NY	454.78
79	Santa Cruz	Santa Cruz	CA	447.03
80	Sacramento	Sacramento	CA	427.14
81	Rensselaer	Albany-Schenectady-Troy	NY	424.19
82	Tuscalossa	Tuscalossa	AL	402.87
83	Kern	Bakersfield	CA	400.74
84	Orange	Orlando	FL	398.00
85	Weber	Salt Lake City-Ogden	UT	393.98
86	Mercer	Trenton	NJ	371.79
87	Sedgwick	Wichita	KS	365.52
88	Santa Clara	San Jose	CA	355.25
89	St. Louis	Duluth-Superior	MN	354.81
90	Allen	Fort Wayne	IN	348.79
91	Champaign	Champaign-Urbana-Rantoul	IL	340.84
92	Monmouth	Long Branch-Asbury Park	NJ	328.84
93	Stanislaus	Modesto	CA	313.42
94	Alameda	San Francisco-Oakland	CA	306.30
95	Cumberland	Fayetteville	NC	301.84
96	Yolo	Sacramento	CA	299.89
97	Bucks	Philadelphia	PA	298.85
98	Pierce	Tacoma	WA	298.78
99	Oneida	Utica-Rome	NY	286.39
100	Berks	Reading	PA	276.10
101	Linn	Cedar Rapids	IA	270.04
102	Volusia	Daytona Beach	FL	267.35
103	Spokane	Spokane	WA	265.20
104	Montgomery	Dayton	OH	263.88
105	San Francisco	San Francisco-Oakland	CA	260.40
106	Onondaga	Syracuse	NY	254.55
107	Lucas	Toledo	OH	250.74
108	Utah	Provo-Orem	UT	249.32
109	Ouachita	Monroe	LA	248.16
110	Lexington	Columbia	SC	246.48
111	Cambria	Johnstown	PA	243.41
112	San Mateo	San Francisco-Oakland	CA	238.14
113	Lancaster	Lincoln	NB	237.02
114	Johnson	Kansas City	KS	229.00
115	Davis	Salt Lake City-Ogden	UT	215.34
116	Schenectady	Albany-Schenectady-Troy	NY	199.24

TABLE 2—CONTINUED

Rank	County	Standard Metropolitan Statistical Area	State	QOLI (1980 dollars)
117	Rock	Janesville-Beloit	WI	196.04
118	Salt Lake	Salt Lake City-Ogden	UT	193.49
119	Woodbury	Sioux City	IA	189.48
		(Mean QOLI = $186)		
120	Davidson	Greensboro-Winston-Salem-High Pt.	NC	178.78
121	Butler	Hamilton-Middletown	OH	168.37
122	Montgomery	Washington	MD	166.10
123	Ingham	Lansing-East Lansing	MI	161.22
124	Ector	Odessa	TX	155.61
125	Sangamon	Springfield	IL	154.34
126	Taylor	Abilene	TX	152.52
127	Erie	Erie	PA	144.24
128	McLean	Bloomington-Normal	IL	140.27
129	Vigo	Terre Haute	IN	137.38
130	San Bernardino	Riverside-San Bernardino-Ontario	CA	135.46
131	Hamilton	Chattanooga	TN	131.69
132	Atlantic	Atlantic City	NJ	128.25
133	Winnebago	Rockford	IL	128.12
134	Benton	Richland-Kennewick-Pasco	WA	113.47
135	St. Joseph	South Bend	IN	108.11
136	Chesterfield	Richmond	VA	107.30
137	Shelby	Memphis	TN	105.46
138	Clackamas	Portland	OR	103.04
139	Lafayette	Lafayette	LA	100.40
140	Smith	Tyler	TX	97.32
141	Douglas	Omaha	NB	94.48
142	Marin	San Francisco-Oakland	CA	92.76
143	Comanche	Lawton	OK	84.37
144	Wyandotte	Kansas City	KS	80.93
145	Ramsey	Minneapolis-St. Paul	MN	79.31
146	Scott	Davenport-Rock Island-Moline	IA	65.83
147	Essex	Newark	NJ	64.67
148	Washington	Portland	OR	58.12
149	Lubbock	Lubbock	TX	57.58
150	Hamilton	Cincinnati	OH	51.79
151	Washtenaw	Ann Arbor	MI	43.74
152	Delaware	Philadelphia	PA	14.18
153	Cabell	Huntington-Ashland	WV	8.32
154	Trumbull	Youngstown-Warren	OH	5.20
155	Dist. of Col.	Washington	DC	3.35
156	Howard	Baltimore	MD	1.90
157	Etowah	Gadsden	AL	1.03
158	King	Seattle-Everett	WA	−7.54
159	Calhoun	Anniston	AL	−10.60
160	Elkhart	Elkhart	IN	−12.50
161	Bexar	San Antonio	TX	−22.07
162	Kenton	Cincinnati	KY	−24.39
163	Rock Island	Davenport-Rock Island-Moline	IL	−24.46
164	Clayton	Atlanta	GA	−25.74
165	Alachua	Gainesville	FL	−26.34
166	Philadelphia	Philadelphia	PA	−28.10
167	Black Hawk	Waterloo-Cedar Falls	IA	−29.70
168	E. Baton Rouge	Baton Rouge	LA	−39.09
169	Brown	Green Bay	WI	−43.73
170	Westchester	New York	NY	−72.12
171	Lake	Chicago	IL	−78.61
172	Kenosha	Kenosha	WI	−85.84
173	Lorain	Lorain-Elyria	OH	−86.76
174	Polk	Lakeland-Winter Haven	FL	−100.69

TABLE 2—CONTINUED

Rank	County	Standard Metropolitan Statistical Area	State	QOLI (1980 dollars)
175	Richland	Mansfield	OH	−104.10
176	Erie	Buffalo	NY	−104.22
177	Chester	Philadelphia	PA	−105.33
178	Outagamie	Appleton-Oshkosh	WI	−107.14
179	Orleans	New Orleans	LA	−118.47
180	Franklin	Columbus	OH	−120.61
181	Travis	Austin	TX	−129.26
182	Prince Geo.'s	Washington	MD	−133.44
183	Union	Newark	NJ	−140.05
184	Cuyahoga	Cleveland	OH	−142.85
185	Kankakee	Kankakee	IL	−143.84
186	DuPage	Chicago	IL	−158.03
187	Yellowstone	Billings	MT	−174.55
188	Mahoning	Youngstown-Warren	OH	−177.13
189	McLennan	Waco	TX	−186.91
190	Yakima	Yakima	WA	−188.39
191	Davidson	Nashville-Davidson	TN	−198.20
192	Duval	Jacksonville	FL	−238.25
193	Bay	Bay City	MI	−247.37
194	Stark	Canton	OH	−250.71
195	St. Clair	St. Louis	IL	−251.01
196	Jefferson	Beaumont-Port Arthur-Orange	TX	−260.20
197	Galveston	Galveston-Texas City	TX	−262.24
198	Hillsborough	Tampa-St. Petersburg	FL	−279.80
199	Marion	Indianapolis	IN	−284.10
200	Jefferson	Louisville	KY	−306.50
201	Macon	Decatur	IL	−308.75
202	Allegheny	Pittsburgh	PA	−308.92
203	Oakland	Detroit	MI	−314.44
204	Middlesex	New Brunswick-Perth Amboy	NJ	−318.20
205	Porter	Gary-Hammond-Each Chicago	IN	−320.47
206	Summit	Akron	OH	−329.96
207	Burlington	Philadelphia	NJ	−346.55
208	Monroe	Toledo	MI	−367.23
209	Dallas	Dallas-Ft. Worth	TX	−372.70
210	Wichita	Wichita Falls	TX	−381.59
211	Contra Costa	San Francisco-Oakland	CA	−386.62
212	Tarrant	Dallas-Ft. Worth	TX	−397.10
213	St. Clair	Detroit	MI	−406.06
214	Harford	Baltimore	MD	−410.29
215	Westmoreland	Pittsburgh	PA	−412.15
216	New York	New York	NY	−414.62
217	Anne Arundel	Baltimore	MD	−430.65
218	Nueces	Corpus Christi	TX	−436.06
219	Bergen	New York	NJ	−465.97
		(Mean QOLI − 1 S.D. = $ − 481)		
220	Baltimore City	Baltimore	MD	−485.32
221	St. Charles	St. Louis	MO	−486.10
222	Hennepin	Minneapolis-St. Paul	MN	−488.20
223	Camden	Philadelphia	NJ	−523.00
224	Saginaw	Saginaw	MI	−537.30
225	Clark	Portland	WA	−547.30
226	Dakota	Minneapolis-St. Paul	MN	−558.10
227	Snohomish	Seattle-Everett	WA	−562.70
228	Allen	Lima	OH	−585.10
229	Jackson	Jackson	MI	−635.30
230	Will	Chicago	IL	−676.10
231	Greene	Dayton	OH	−681.30
232	Niagara	Buffalo	NY	−682.70

TABLE 2—CONTINUED

Rank	County	Standard Metropolitan Statistical Area	State	QOLI (1980 dollars)
233	Calhoun	Battle Creek	MI	−701.10
234	Denton	Dallas-Ft. Worth	TX	−709.90
235	Peoria	Peoria	IL	−758.80
236	Rockland	New York	NY	−795.50
237	Cameron	Brownsville-Harlingen-San Benito	TX	−795.70
238	Medina	Cleveland	OH	−823.30
239	Hidalgo	McAllen-Pharr-Edinburg	TX	−823.80
240	St. Louis	St. Louis	MO	−875.30
241	Harris	Houston	TX	−916.30
242	Jefferson	St. Louis	MO	−918.30
243	Washington	Minneapolis-St. Paul	MN	−920.20
244	Kent	Grand Rapids	MI	−950.90
245	Kalamazoo	Kalamazoo-Portage	MI	−976.30
246	Cook	Chicago	IL	−979.10
247	Genesee	Flint	MI	−1018.50
248	Macomb	Detroit	MI	−1024.10
		(Mean QOLI − 2 S.D. = $ − 1148)		
249	Wayne	Detroit	MI	−1267.50
250	Brazoria	Houston	TX	−1403.50
251	Jefferson	Birmingham	AL	−1539.30
252	Waukesha	Milwaukee	WI	−1791.50
253	St. Louis City	St. Louis	MO	−1856.70

TABLE 3—INTRAURBAN VARIATION IN QUALITY OF LIFE

Standard Metropolitan Statistical Area	Counties	Range of QOLI (dollars)	Average QOLI[a] (dollars)	County QOLI (dollars)	County Rank[b]
Milwaukee, WI		$2350	$29		
	Milwaukee			$558	68
	Waukesha			−1792	252
St. Louis, MO-IL		1606	−990		
	St. Clair, IL			−251	195
	St. Charles, MO			−486	221
	St. Louis, MO			−875	240
	Jefferson, MO			−918	242
	St. Louis City, MO			−1857	253
Denver-Boulder, CO		1488	1198		
	Arapahoe			2097	3
	Boulder			1320	12
	Denver			1295	14
	Adams			727	53
	Jefferson			609	62
Norfolk-Virginia-Beach-Portsmouth, VA-NC		1138	1542		
	Norfolk City			2106	2
	Virginia Beach City			968	28
Philadelphia, PA-NJ		1057	9		
	Montgomery, PA			534	71
	Bucks, PA			299	97
	Delaware, PA			14	152
	Philadelphia, PA			−28	166
	Chester, PA			−105	177
	Camden, NJ			−523	223

TABLE 3—CONTINUED

Standard Metropolitan Statistical Area	Counties	Range of QOLI (dollars)	Average QOLI[a] (dollars)	County QOLI (dollars)	County Rank[b]
Minneapolis-St. Paul, MN		999	−372		
	Ramsey			79	145
	Hennepin			−488	222
	Dakota			−558	226
	Washington			−920	243
Detroit, MI		954	−968		
	Oakland			−314	203
	St. Clair			−406	213
	Macomb			−1024	248
	Wayne			−1268	249
Richmond, VA		953	686		
	Richmond City			1060	20
	Chesterfield			107	136
Dayton, OH		945	89		
	Montgomery			264	104
	Greene			−681	231
Chicago, IL		900	−823		
	Lake			−79	171
	DuPage			−158	186
	Will			−676	230
	Cook			−979	246

[a] The average QOLI values are population-weighted averages of the county QOLI values.
[b] The county rank is taken from Table 2.

TABLE 4—DIFFERENCES IN MEASURED QUALITY OF LIFE BY AMENITY BUNDLE, HOURS WORKED AND POPULATION GROWTH

Rank	County	Standard Metropolitan Statistical Area	State	QOLI	QOLI Rank: Climate[a]	QOLI Rank: Environ. Quality[b]	QOLI Rank: Urban Condit.[c]	QOLI Rank: Low Hrs. Worked[d]	QOLI Rank: High Hrs. Worked[e]	Population Growth[f]
1	Pueblo	Pueblo	CO	$3289	9	221	1	1	1	0
2	Norfolk City	Norfolk-VA Beach-Portsmouth	VA	2106	35	13	5	17	2	−
3	Arapahoe	Denver-Boulder	CO	2097	29	168	3	2	9	+
4	Bibb	Macon	GA	1600	102	7	4	40	6	0
5	Washoe	Reno	NV	1575	3	171	130	6	26	+
6	Broome	Binghamton	NY	1486	239	138	2	10	27	−
7	Hampton City	Newport News-Hampton	VA	1445	35	3	48	31	12	0
8	Sarasota	Sarasota	FL	1431	47	2	26	128	3	+
9	Palm Beach	West Palm Beach-Boca Raton	FL	1423	10	40	102	69	5	+
10	Pima	Tucson	AZ	1342	4	191	151	58	10	+
11	Broward	Ft. Lauderdale-Hollywood	FL	1327	20	188	33	106	4	+
12	Boulder	Denver-Boulder	CO	1319	29	201	28	12	47	+
13	Larimer	Ft. Collins	CO	1298	29	170	50	20	32	+
14	Denver	Denver-Boulder	CO	1295	29	207	29	22	29	−
15	Charleston	Charleston-N. Charleston	SC	1280	14	38	110	41	19	0
*										
*										
*										
120	Davidson	Greensboro-Win. Sal.-High Pt.	NC	$179	83	21	198	174	90	0
121	Butler	Hamilton-Middletown	OH	168	127	195	92	188	82	0
122	Montgomery	Washington	MD	166	91	95	183	139	117	0
123	Ingham	Lansing-E. Lansing	MI	161	249	20	16	125	128	0

TABLE 4—CONTINUED

Rank	County	Standard Metropolitan Statistical Area	State	QOLI	QOLI Rank: Climate[a]	QOLI Rank: Environ. Quality[b]	QOLI Rank: Urban Condit.[c]	QOLI Rank: Low Hrs. Worked[d]	QOLI Rank: High Hrs. Worked[e]	Population Growth[f]
124	Ector	Odessa	TX	156	85	133	173	119	130	0
125	Sangamon	Springfield	IL	154	199	25	76	99	146	0
126	Taylor	Abilene	TX	153	120	105	154	170	105	0
127	Erie	Erie	PA	144	174	216	35	67	169	0
128	McLean	Bloomington-Normal	IL	140	206	32	73	154	112	0
129	Vigo	Terre Haute	IN	137	189	24	86	180	98	–
130	San Bernardino	Riverside-San Bernardino-Ontario	CA	135	54	161	226	59	188	0
131	Hamilton	Chattanooga	TN	132	73	225	135	241	46	0
132	Atlantic	Atlantic City	NJ	128	65	197	182	29	209	0
133	Winnebago	Rockford	IL	128	201	136	61	135	132	0
134	Benton	Richland-Kennewick-Pasco	WA	113	147	140	125	142	133	+
*										
*										
*										
239	Hidalgo	McAllen-Pharr-Edinburg	TX	−824	253	83	75	252	155	+
240	St. Louis	St. Louis	MO	−875	195	193	227	226	230	0
241	Harris	Houston	TX	−916	181	250	100	245	218	+
242	Jefferson	St. Louis	MO	−918	195	164	237	235	226	+
243	Washington	Minneapolis-St. Paul	MN	−920	220	125	235	227	237	+
244	Kent	Grand Rapids	MI	−951	251	238	78	160	245	0
245	Kalamazoo	Kalamazoo-Portage	MI	−976	251	189	165	197	242	0
246	Cook	Chicago	IL	−979	135	251	168	150	248	–
247	Genesee	Flint	MI	−1018	238	200	212	201	243	0
248	Macomb	Detroit	MI	−1024	229	173	231	183	246	0
249	Wayne	Detroit	MI	−1268	124	247	242	243	241	–
250	Brazoria	Houston	TX	−1404	156	252	211	231	250	+
251	Jefferson	Birmingham	AL	−1539	241	248	188	253	238	0
252	Waukesha	Milwaukee	WI	−1792	244	253	113	177	253	0
253	St. Louis City	St. Louis	MO	−1857	195	184	253	251	252	–
Rank Correlation with QOLI (Table 2)			r	1.00	0.63	0.21	0.48	0.63	0.83	0.16

[a] The QOLI-climate ranking is based on QOLI values which are calculated using the full implicit prices shown in Table 1 and the subset of amenities which includes only: precipitation, humidity, heating degree days, cooling degree days, wind speed, sunshine, and coast.

[b] The QOLI-environmental quality ranking is based on the subset of amenities which includes only: visibility, total suspended particulates, NPDES effluent dischargers, landfill waste, superfund sites, and treatment, storage, and disposal sites.

[c] The QOLI-urban conditions ranking is based on the subset of amenities which includes only: central city, violent crime, and teacher-pupil ratio.

[d] The QOLI-low hours worked is based on QOLI values which are calculated as described in note c to Table 1 except the annual number of hours worked is halved.

[e] The QOLI-low hours worked is based on QOLI values which are calculated as described in note c to Table 1 except the annual number of hours worked is doubled.

[f] Growth is the percentage change in county population between 1970 and 1980. A plus is assigned to counties with growth greater than twice the average (> 33.4 percent) and a minus is assigned to counties with growth less than zero. A zero is assigned to counties with growth between zero and twice the average. The rank correlation reported is for QOLI and population growth.

ranking is .63, for environmental quality it is .21, and for urban conditions it is .48. Thus, the ranking for households who value only a subset of amenities can be quite different. These rankings for amenity subsets are also useful for illustrating which factors contribute to an individual county's high or low ranking. Pueblo, Colorado, the top-ranked county, has a high ranking for climate and a very low one for environmental quality. But it has the top ranking for urban conditions (in particular, the teacher-pupil ratio), which is an important contributor to its top overall ranking. Ingham (Lansing), Michigan, has very high rankings for environmental quality and urban conditions, but is almost on the bottom for climate, giving it a middle overall ranking. Thus, a household which did not consider climate as a component of quality of life would find Ingham County an attractive place to live. St. Louis City is at the bottom of the ranking. While its ranking is fairly low for all three subsets, its ranking for urban conditions is at the bottom. The high violent crime rate is an important contributor to the bottom ranking for urban conditions and to its overall ranking.

Finally, in order to address the possibility of disequilibrium influences, the last column of Table 4 shows the population growth the county experienced between 1970 and 1980. A plus sign indicates that the county grew by more than twice the average (>33.4 percent), a minus sign is given to counties which experienced population loss between 1970 and 1980, while a zero indicates population growth between zero and 33.4 percent. There is a small but significant positive correlation (.16) between population change and the standard QOLI. Eight out of the 15 top counties are fast (+) growing, 13 out of 15 middle counties have had slow (0) growth, while among the bottom 15, there is a mix of fast and slow growth, and population losses. Overall, it does not appear that disequilibrium influences are overwhelming, and if anything, they reinforce the standard equilibrium-based QOLI.

V. Conclusions

This paper provides new quality of life rankings both across and within urban areas. The estimation is based on a national hedonic model which incorporates variation in both wages and housing expenditures. The wage and housing-expenditure equations are estimated using micro data from the 1980 Census and aggregate data on various amenities from other sources.

A quality of life index is constructed using preference-based weights, which are the amenity values derived from the hedonic estimation. The index is calculated for 253 counties, many of which are located in the same SMSA. The difference between the top- and bottom-ranked counties is estimated to be valued at \$5146 per household per year. The highest ranked counties tend to be connected with small- and medium-sized cities in the Sun Belt and Colorado. The bottom of the ranking is dominated by medium to large northern cities. Substantial intracity variation in measured quality of life is found. For example, \$1606 per household per year is the difference across counties in the St. Louis SMSA.

The results indicate that compensation for location-specific, nontraded amenities takes place in both the labor and housing market and that the amount is substantial. Quality of life is clearly one factor considered in location decisions along with other factors such as job availability. Indexes which include a set of factors different from ours may yield different rankings for different purposes. Nonetheless, our ranking is based on amenities which are an important component of economic well-being. The valuations of these nontraded amenities yield new quality of life comparisons across and within a wide variety of American urban areas.

APPENDIX

Comparative Static Results. Equilibrium wage, rent, and city-size differentials are solved for using the total differential of equation system (S1). This differential system is

$$\text{(A1)}\quad \begin{vmatrix} -v^1_{a_1} \\ -c^1_{a_1} \\ 0 \\ 0 \\ -N^1_{a_1} \end{vmatrix} = \begin{vmatrix} v^1_{w_1} & v^1_{r_1} & 0 & 0 & 0 \\ c^1_{w_1} & 0 & 0 & 0 & c^1_N \\ 0 & 0 & v^2_{w_2} & v^2_{r_2} & 0 \\ 0 & 0 & c^2_{w_2} & 0 & c^2_N \\ N^1_{w_1} & N^1_{r_1} & N^2_{w_2} & N^2_{r_2} & -1 \end{vmatrix} \begin{vmatrix} dw_1/da_1 \\ dr_1/da_1 \\ dw_2/da_1 \\ dr_2/da_1 \\ dN/da_1 \end{vmatrix} = A \begin{vmatrix} dw_1/da_1 \\ dr_1/da_1 \\ dw_2/da_1 \\ dr_2/da_1 \\ dN/da_1 \end{vmatrix}.$$

The determinate of A is

$$\text{(A2)}\quad D = -v_{r_1}^1 c_N^1 c_{w_2}^2 v_{r_2}^2 \left[N_{w_1}^1 - \left(v_{w_1}^1 / v_{r_1}^1\right) N_{r_1}^1\right] - v_{r_2}^2 c_N^2 v_{r_1}^1 c_{w_1}^1 \left[N_{w_2}^2 - \left(v_{w_2}^2 / v_{r_2}^2\right) N_{r_2}^2\right] - v_{r_1}^1 v_{r_2}^2 c_{w_2}^2 c_{w_1}^1 .$$

D is negative if city size, N, is Walrasian stable with respect to a change in wages in county k, $k \in \{1,2\}$. The adjoint of A is

$$\text{(A3)}\quad C' = \begin{vmatrix} c_N^1 N_{r_1}^1 c_{w_2}^2 v_{r_2}^2 & v_{r_1}^1 v_{r_2}^2 \beta_2 & v_{r_1}^1 c_N^1 c_{w_2}^2 N_{r_2}^2 & v_{r_1}^1 v_{r_2}^2 c_N^1 \alpha_2 & -v_{r_1}^1 v_{r_2}^2 c_N^1 c_{w_2}^2 \\ v_{r_2}^2\left(c_{w_1}^1 \beta_2 - c_{w_2}^2 c_N^1 N_{w_1}^1\right) & -v_{r_1}^1 v_{w_1}^1 \beta_2 & -v_{w_1}^1 c_N^1 c_{w_2}^2 N_{r_2}^2 & -v_{w_1}^1 v_{r_2}^2 c_N^1 \alpha_2 & v_{w_1}^1 c_{w_2}^2 v_{r_2}^2 c_N^1 \\ c_{w_1}^1 N_{r_1}^1 c_N^2 v_{r_2}^2 & v_{r_1}^1 v_{r_2}^2 c_N^2 \alpha_1 & v_{r_1}^1 c_{w_1}^1 c_N^2 N_{r_2}^2 & v_{r_1}^1 v_{r_2}^2 \beta_1 & -v_{r_1}^1 c_{w_1}^1 v_{r_2}^2 c_N^2 \\ -c_{w_1}^1 N_{r_1}^1 c_N^2 v_{w_2}^2 & -v_{r_1}^1 v_{w_2}^2 c_N^2 \alpha_1 & -v_{r_1}^1\left(c_{w_2}^2 \beta_1 - c_{w_1}^1 c_N^2 N_{w_2}^2\right) & -v_{r_1}^1 v_{w_2}^2 \beta_1 & v_{r_1}^1 c_{w_1}^1 v_{w_2}^2 c_N^2 \\ -c_{w_1}^1 N_{r_1}^1 c_{w_2}^2 v_{r_2}^2 & -v_{r_1}^1 v_{r_2}^2 c_{w_2}^2 \alpha_1 & -v_{r_1}^1 c_{w_1}^1 c_{w_2}^2 N_{r_2}^2 & -v_{r_1}^1 v_{r_2}^2 c_{w_1}^1 \alpha_2 & v_{r_1}^1 c_{w_1}^1 v_{r_2}^2 c_{w_2}^2 \end{vmatrix}$$

where $\alpha_j = N_{w_j}^j - (v_{wj}^j / v_{rj}^j) N_{rj}^j$ and $\beta_j = c_N^j[(v_{wj}^j / v_{rj}^j) N_{r_j}^j - N_{w_j}^j] - c_{w_j}^j$, $j \in \{1,2\}$. The term α_j is positive if the supply of labor to county j, $N_{w_j}^j$ is positively sloped with respect to change in the wage in county j. The term β_j is negative if the supply of labor to county j is Walrasian stable with respect to a change in wages in county j. $N_{w_j}^j$ is nonpositive if land is not an inferior good. $N_{r_j}^j$ is nonnegative if land is not a Giffen good.

The solution to the differential system is

$$\text{(A4)}\quad \begin{vmatrix} dw_1/da_1 \\ dr_1/da_1 \\ dw_2/da_1 \\ dr_2/da_1 \\ dN/da_1 \end{vmatrix} = (C'/D) \begin{vmatrix} -v_{a_1}^1 \\ -c_{a_1}^1 \\ 0 \\ 0 \\ -N_{a_1}^1 \end{vmatrix}$$

$$= \begin{vmatrix} -v_{a_1}^1 c_N^1 c_{w_2}^2 v_{r_2}^2 N_{r_1}^1 - c_{a_1}^1 v_{r_1}^1 v_{r_2}^2 \beta_2 + N_{a_1}^1 v_{r_1}^1 v_{r_2}^2 c_N^1 c_{w_2}^2 \\ -v_{a_1}^1 v_{r_2}^2\left(c_{w_1}^1 \beta_2 - c_{w_2}^2 c_N^1 N_{w_1}^1\right) + c_{a_1}^1 v_{w_1}^1 v_{r_1}^1 \beta_2 - N_{a_1}^1 v_{w_1}^1 v_{r_2}^2 c_N^1 c_{w_2}^2 \\ -v_{a_1}^1 c_N^2 c_{w_1}^1 v_{r_2}^2 N_{r_1}^1 - c_{a_1}^1 v_{r_1}^1 v_{r_2}^2 c_N^2 \alpha_1 + N_{a_1}^1 c_N^2 c_{w_1}^1 v_{r_1}^1 v_{r_2}^2 \\ v_{a_1}^1 c_N^2 c_{w_1}^1 v_{w_2}^2 N_{r_1}^1 + c_{a_1}^1 v_{r_1}^1 v_{w_2}^2 c_N^2 \alpha_1 - N_{a_1}^1 c_N^2 c_{w_1}^1 v_{w_2}^2 v_{r_1}^1 \\ v_{a_1}^1 c_{w_2}^2 v_{r_2}^2 c_{w_1}^1 N_{r_1}^1 + c_{a_1}^1 v_{r_1}^1 v_{r_2}^2 c_{w_2}^2 \alpha_1 - N_{a_1}^1 v_{r_2}^2 c_{w_2}^2 v_{r_1}^1 c_{w_1}^1 \end{vmatrix} /D.$$

Multiplying the cross-county rent differential, dr_2/da_1, by $v_{r_2}^2 / v_{r_2}^2$ and factoring out the term $-v_{w_2}^2$ results in

$$\text{(A5)}\quad dr_2/da_1 = \left(-v_{w_2}^2 / v_{r_2}^2\right)\left(-v_{a_1}^1 c_N^2 c_{w_1}^1 v_{r_2}^2 N_{r_1}^1 - c_{a_1}^1 v_{r_1}^1 v_{r_2}^2 c_N^2 \alpha_1 + N_{a_1}^1 c_N^2 c_{w_1}^1 v_{r_1}^1 v_{r_1}^2\right)/D = (1/q_2)(dw_2/da_1).$$

A result analogous to (A5) can be derived for dr_1/da_2.

For the case described in Section I, Part B, of the text, the signs of the differentials hold under the assumption that $N_{a_k}^k = 0$, $N_{w_k}^k \leq 0$, and $N_{r_k}^k \geq 0$. An additively separable utility function is sufficient for N_{a_k} to equal zero. If land is not inferior and land is not a Giffen good, then $N_{w_k}^k \leq 0$ and $N_{r_k}^k \geq 0$.

REFERENCES

Cobb, Steven, "The Impact of Site Characteristics on Housing Cost Estimates," *Journal of Urban Economics*, January 1984, *15*, 26–45.

Executive Office of the President, *Environmental Quality*, 1982, 13th Annual Report of the Council of Environmental Quality, Washington, DC: USGPO, 1983, Appendix C.

Graves, Philip E., "Migration with a Composite Amenity: The Role of Rents," *Journal of Regional Science*, November 1983, *23*, 541–46.

Hoehn, John P., Berger, Mark C. and Blomquist, Glenn C., "A Hedonic Model of Interregional Wages, Rents and Amenity Values," *Journal of Regional Science*, November 1987, *27*, 605–20.

Johnston, J. *Econometric Methods*, 3rd ed., New York: McGraw-Hill, 1984.

Kokkelenberg, Edward C. and Sockell, Donna R., "Union Membership in the United States, 1973–1981," *Industrial and Labor Relations Review*, July 1985, *38*, 497–543.

Krumm, Ronald J., "Neighborhood Amenities: An Economic Analysis," *Journal of Urban Economics*, March 1980, *7*, 208–24.

Linneman, Peter, "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," *Journal of Urban Economics*, July 1980, *8*, 47–68.

Liu, Ben-Cheih, *Quality of Life Indicators in U.S. Metropolitan Areas*, New York: Praeger Publishers, 1976.

Peiser, Richard B. and Smith, Lawrence B., "Homeownership Returns, Tenure Choice and Inflation," *American Real Estate and Urban Economics Journal*, Winter 1985, *13*, 343–60.

Roback, Jennifer, "Wages, Rents and Amenities: Differences Among Workers and Regions, *Economic Inquiry*, January 1988, *26*, forthcoming.

______, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, December 1982, *90*, 1257–78.

Rosen, Sherwin, "Wage-Based Indexes of Urban Quality of Life," in Peter Mieszkowski and Mahlon Straszheim, eds., *Current Issues in Urban Economics*, Baltimore: Johns Hopkins University Press, 1979.

Smith, V. Kerry, "The Role of Site and Job Characteristics in Hedonic Wage Models," *Journal of Urban Economics*, May 1983, *13*, 296–321.

Tolley, George S., "The Welfare Economics of City Bigness," *Journal of Urban Economics*, July 1974, *1*, 324–45.

Trijonis, John and Shapland, Dawn, "Existing Visibility Levels in the United States," Technology Services Corp., Santa Fe, NM, for USEPA (EPA-450 15-79-010), 1979.

U.S. Bureau of Census, *Census of Government, 1982*, Washington, DC: USGPO, 1984, Vol. 3, Table 14, and Vol. 4, Table 8.

U.S. Bureau of Census, *Census of Population and Housing, 1980: Public-Use Microdata Samples Technical Documentation* prepared by the Data User Services Division, Washington, DC: The Bureau, 1983.

U.S. Department of Commerce, *Comparative Climatic Data*, Asheville, NC: National Climatic Data Center, August 1983.

U.S. Environmental Protection Agency, *RCRA Application for Hazardous Waste Permit Tape*, Washington, DC: USEPA, 1984.

U.S. Environmental Protection Agency, *SAROAD Storage and Retrieval of Aerometric Data Information*, Research Triangle Park, USEPA, 1979.

U.S. Federal Bureau of Investigation, *FBI Uniform Crime Reports for the United States*, Washington, DC: USGPO, 1980.

The Division of Markets Is Limited by the Extent of Liquidity (Spatial Competition with Externalities)

By NICHOLAS ECONOMIDES AND ALOYSIUS SIOW*

Liquidity considerations will limit the number of markets in a competitive economy. Welfare implications are ambiguous. Since liquidity is a positive externality, there may be too little liquidity per market at a noncooperative equilibrium and too many markets compared to the surplus-maximizing market structure. But liquidity is also self-reinforcing. Given an existing equilibrium, new markets may not open because nobody wants to use a new market with low liquidity. There may be too few markets to achieve efficiency and new markets will not open. A nondiscriminating monopolist will operate smaller and more numerous markets compared to optimality as well as to the equilibrium of independent auctioneers.

How many markets are there in a competitive economy? In the standard general equilibrium model (for example, Gerard Debreu, 1959), where there is no trading friction, there are as many markets as there are commodities. When there are trading frictions due to technological or demographic constraints, the number of markets may be less than the number of commodities (for example, David Cass and Karl Shell, 1983; Peter Diamond, 1982; Robert Townsend, 1983).

This paper studies the role of lack of liquidity as an endogenous trading friction in limiting the number of markets in a competitive economy. In many markets, the variance of (competitive) price fluctuations is negatively correlated with the volume of trade in that market. We define a market as having *high liquidity* when the volume of trade is high and the corresponding variance of price is low.[1] The problem of liquidity is most apparent in financial markets. For example, liquidity is a particularly important factor in determining the success of futures contracts. The futures market for any asset has only a small number of maturity dates. In principle, many more maturity dates may be admitted. However, if there were many maturity dates, the market at each maturity date would be thin. Market participants may face large competitive price fluctuations arising only from the thinness of the markets. Traders may prefer fewer maturity dates so that liquidity is enhanced in the remaining markets, even though they will have fewer maturity dates to choose from. Thus, there is a fundamental tradeoff between liquidity and the number of markets.[2] In a geographic context, the historical development of spa-

---

*Department of Economics, Columbia University, New York, NY 10027. We thank Ken Arrow, Ralph Braid, Tina Fine, Douglas Holtz-Eakin, Kel Lancaster, Martin Osborne, Carolyn Pitchik, Susan Rose-Ackerman, Robert Townsend, Chuck Wilson, Mike Woodford, and members of workshops at Columbia University, New York University, and the IMSSS summer workshop at Stanford for their comments. We are also grateful to the referees for their patience and comments. We acknowledge support from the National Science Foundation under grant nos. SES-8408905 and SES-8411396, and the Columbia Council for Research in the Social Sciences.

[1] Our definition ignores the speed at which sales can be consummated. For example, Steven Lippman and John McCall (1986) define an asset as liquid "if it can be sold quickly and at a predictable price." Their paper contains a comprehensive discussion of the attributes of liquidity that are important to a seller of an asset.

[2] See Dennis Carlton, 1984; Kenneth Garbade and William Silber, 1979; Lester Telser, 1981, and the references therein for discussions of liquidity and the success of futures markets. Deborah Black (1985) contains empirical evidence on the same subject.

tially separated towns may be attributed to the tradeoff that farmers faced between liquidity in the trading place and costs of transporting the commodities to the market.

This paper was motivated by Diamond's search models (for example, Peter Diamond, 1982, 1984; Dale Mortensen, 1976; Aloysius Siow, 1982). Many issues addressed in this paper were raised by him, and our analysis complements his work. Our formal model uses a spatial location framework which is related to work by Townsend (1983, 1984). Townsend's models address some of the same concerns as ours. The main differences between our work and his are that our model is analytically more tractable and we use the Nash equilibrium as our main solution concept. We briefly consider the core as an alternative solution concept in the final section of the paper.

In Section I we show that each agent in the economy, faced with uncertain endowments, prefers to trade in a spot market with high rather than low liquidity. Liquidity at a market can only be increased by increasing the number of traders at that market. The traders in our economy are spatially separated so that as more traders go to a particular market, traders are coming from farther away. The farther a trader has to travel to a market, the larger is his transportation cost. Because of this fundamental tradeoff between liquidity and transportation costs, all agents in the economy will not go to the same market.

This paper considers two kinds of competitive market structures. The first assumes that traders may participate in markets without charge (as in standard Walrasian markets). The second structure assumes that it is costly to operate a market, which means that the "auctioneer" must be paid for providing market services. In Section II we establish and characterize the noncooperative symmetric equilibria for the economy with free market services. A basic positive result from this section is that even without fixed cost, liquidity considerations will limit the number of markets in a competitive economy. The welfare implications of the competitive division of markets in this economy are ambigious. Since liquidity is a positive externality, there may be too little liquidity at equilibrium because each agent acts only in his own self-interest. In this case there are too many markets to be efficient. On the other hand, liquidity is self-reinforcing. Given an existing market structure, new markets may find it impossible to open because nobody wants to use a new market with low liquidity. There may be fewer markets than is necessary for efficiency, and yet new markets will not open.

In Section III we make it costly to operate a market. Each market must be operated by a *market maker* whose opportunity cost is the expected utility that he can get as a trader. Arbitrage between being a market maker or a trader reduces the number of equilibria relative to the noncooperative equilibrium with free market services. However, competition between market makers for customers is not sufficient to internalize the externality caused by liquidity. This result runs counter to that of Frank Knight (1924) who suggested that profit-maximizing ownership of a congested facility leads to efficient pricing.[3]

In Section IV we study the market structure when all markets are organized and run by a monopoly exchange. We show that the monopolist will overcrowd the space with small markets.

Liquidity in our model is not tied to the spot market specification in the economy. In Section V we show that the same issues arise when the spot markets are replaced by state-contingent claims between agents at a specific market location. State-contingent claims markets cannot reduce the intrinsic uncertainty of any specific market location. Only the addition of traders at a specific market location may reduce the uncertainty at that location. Since contingent claims markets cannot reduce transportation costs, the same issues remain. In Section V we also briefly consider the concept of the core as an alternative equilibrium concept. Final remarks are also in this section.

---

[3] Exceptions to Knight's result have been noted elsewhere. For references and a study of duopoly pricing of congested facilities, see Ralph Braid (1986).

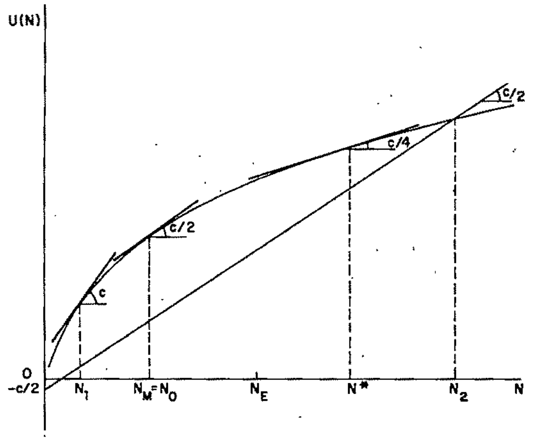


FIGURE 1. LOCATIONS OF AGENTS 1,2,..., N AND MARKETS m_1 AND m_2

I. The Model

Let consumers be located at unit distances apart on the real line (see Figure 1). Each consumer receives at his location a stochastic endowment of two goods x and y. The endowment consists of commodity vector (1,0) with probability $1-\theta$ and of commodity vector (0,1) with probability θ. Consumers have identical preferences over the lotteries of goods x and y received at their location. If a trader goes to another location to exchange goods, the utility of the bundle of goods at the new location is reduced by the disutility from traveling that a trader has to incur to reach the new location.

Consumers, who are expected utility maximizers, can meet each other at any location to exchange goods. The locations at which they meet are called markets.[4] The decision whether to participate in a market is made before the endowment vector is realized. If the consumer does not participate, he consumes his endowment and receives a utility of zero. If he participates in a market, his endowment is realized after he arrives at the market. After the endowments are realized, all consumers in the same market may trade with one another in a competitive spot market for the two goods. Since he has to make his participation decision before he learns of his endowment, a trader evaluates his return from participation by calculating his expected utility from participation. We will show that *ex ante* the expected price at a market is independent of the size of the market, and that the variance of price is lower at a market of high liquidity (with many traders). Thus, a consumer prefers to be in a market of high liquidity.

A consumer incurs disutility from traveling to the market, so that, *ceteris paribus*, he prefers a market closer to his location on the line. Large participation in a market requires that some consumers travel a great distance. Thus, there is a tradeoff between market liquidity and distance between markets. This tradeoff determines the equilibrium distribution of markets in the economy.

When endowments are realized, a consumer may receive one unit of x and zero y, in which case we call him of type 1. Alternatively, a consumer may receive zero units of x and one unit of y, and he is called of type 2. Let a market of N participants consist of X traders of type 1 and Y traders of type 2. Let $k = Y/N$ be the proportion of type 2 traders. A realized market can be described by the pair (k, N). Y is distributed binomially (N, θ). It follows that $E(k|N) = E(Y/N|N) = \theta$.

Assume that exchange in every realized market is Walrasian. Aggregate supply and demand for each commodity are proportional to N, so that the equilibrium price is independent of N.[5] Let $V_1(k)$ (respectively,

---

[4] The spatial separation of traders on a line has an obvious geographic interpretation. It can also be thought of as measuring time, in the framework of trading in futures' contracts. Then the original location of a trader is interpreted as the date when a trader will receive an uncertain endowment vector. Positions where traders meet (markets) are interpreted as the maturity dates of futures' contracts. Travel costs then represent the utility lost in storage costs covering the time interval between the revelation of endowment and the fulfillment of the futures contract.

[5] Call $x_i(P)$ the demand of a type i trader, where P is the price of y relative to x. Market clearing is defined by $Xx_1(P)+Yx_2(P)=X \Leftrightarrow. (1-k)x_1(P)+kx_2(P)=1-k$, which defines a price $P(k)$ independent of N, the size of the market.

$V_2(k)$) denote the indirect utility of a trader of type 1 (respectively of type 2) in a market (k, N). Conditional on a particular value of k, the expected utility of a trader who does not know his type is

$$(1) \quad W(k) \equiv (1-k)V_1(k) + kV_2(k).$$

Since k is a random variable, the unconditional expected utility of a trader in a market of N traders $(N \geq 1)$ is

$$(2) \qquad U(N) = E\{W(k)|N\}.$$

For large N, $W(k)$ can be approximated by a Taylor expansion up to the second order around $k = \theta$:

$$\begin{aligned}(3) \quad U(N) &= E\{W(k)|N\} \\ &\simeq E\{[W(\theta) + (k-\theta)W'(\theta) \\ &\qquad + (k-\theta)^2 W''(\theta)/2]|N\} \\ &= W(\theta) + W''(\theta)\theta(1-\theta)/(2N),\end{aligned}$$

since $E(k|N) = \theta$, $E\{(k-\theta)^2|N\} = \mathrm{var}(k) = \theta(1-\theta)/N$.[6] For the remainder of this paper, we assume that consumers have either Cobb-Douglas or CES utility functions. As shown in Appendix A, this implies that $W(k)$ is concave and $W''(\theta) < 0$.[7]

$U(N)$, the benefit of a trader from participating in a market of N traders, is an increasing and concave function of N.[8] A trader prefers to be in a larger market, but the marginal advantage decreases as the market becomes larger. Traders prefer larger markets because they provide higher liquidity through lower price variance. However, a trader has to travel from his location to the market to be able to participate in it. The disutility of travel must be subtracted from $U(N)$ to determine the net benefit of participation. We assume that the disutility of travel is linear in the distance traveled at rate c per unit of distance traveled. Thus, a trader participating in a market of N traders at distance α from his original location has net benefit $U(N) - c\alpha$.

II. Noncooperative Equilibria with Free Market Services

We can now look for an equilibrium with free market services. Let agents $j = 1, \ldots$ be located on a real line at consecutive positions one unit distance apart. A strategy of an agent is his choice of final location on the line. He can stay at his initial location and consume his endowment (earning zero utility), or travel to another location on the line in order to trade with other agents. The utility an agent attains by participating in a market depends on how many other agents participate in the same market.

A market structure is a noncooperative equilibrium of the game of market participation. At a noncooperative equilibrium, every agent chooses to travel to a location that maximizes his expected utility under the expectation that all other agents will not change their decision with respect to their market affiliation.

There are many equilibria in this game, including quite unreasonable ones. For example, there is an equilibrium in which everybody stays home because everybody expects all others to stay home. We restrict our attention to symmetric equilibria where every trader participates in a market.

DEFINITION: *At a symmetric equilibrium the distance between neighboring markets is N, and there are N participants per market. Each trader chooses to travel to a location that maximizes his expected utility assuming that all other traders will not change the locations to which they travel.*

---

[6] Since Y is binomial (N, θ), $\mathrm{var}(k) = \mathrm{var}(Y/N) = (1/N^2)\mathrm{var}(Y) = \theta(1-\theta)/N$. The approximation of equation (2) is good for large N.

[7] The concavity of $W(k)$ can also be derived by allowing agents to participate in location-specific state contingent markets, where we only assume agents have concave utility functions. See Section V.

[8] $dU(N)/dN = -W''(\theta)\theta(1-\theta)/(2N^2) > 0$, $d^2U/dN^2 = W''(\theta)\theta(1-\theta)/N^3 < 0$.

In a symmetric configuration with N traders in each market, we have to make sure that traders prefer to go to the market closest to them. With reference to Figure 1, let market m_1 be located in $1/2$ and market m_2 be located at $N+1/2$. Between these two markets, there is a consumer initially located at every integer from 1 to N. Consider the marginal consumer of market m_1, at distance $(N/2-1/2)$ from m_1 and distance $(N/2+1/2)$ from m_2. He weakly prefers to participate in market m_1 rather than in m_2 if

$$U(N)-c(N-1)/2 \geq U(N+1)-c(N+1)/2,$$

or equivalently

$$U(N+1)-U(N) \leq c.$$

This is approximated by

$$U'(N) \leq c. \tag{4}$$

Let N_1 be the minimal N obeying this inequality,[9] so that inequality (4) is satisfied for all $N \geq N_1$ (see Figure 2). Given that the marginal consumer prefers to go to market m_1 rather than m_2 (under the expectation that all consumers between him and m_1 to go to m_1), any other consumer i closer to m_1 also prefers to go to m_1 rather than m_2 (under the expectation that all other consumers between $1/2$ and $(N/2-1/2)$ go to m_1). This comes directly from the concavity of $U(N)$ in N.

From inequality (4) we see that a necessary condition for the existence of a symmetric equilibrium is that each market has at least N_1 participants. Thin markets located close to one another cannot constitute an equilibrium.

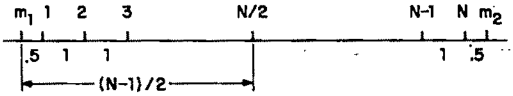


Figure 2. Markets of Sizes N in (N_1, N_2) Are Noncooperative Equilibria. N^* Is the Optimal Market Size. N_M Is the Monopoly Market Size. N_E Is the Equilibrium Market Size with Independent Market Makers

For the equilibrium to exist, we also require that the marginal trader at distance $(N-1)/2$ from m_1, be better off by participating in the market rather than staying home, that is,

$$U(N)-c(N-1)/2 \geq 0. \tag{5}$$

Let N_2 be the solution of (5) as equality. All $N \leq N_2$ satisfy inequality (5). Markets cannot be too large because traders from afar do not wish to participate. Inequalities (4) and (5) are necessary and sufficient for the existence of a symmetric equilibrium. Thus, all markets of sizes N in $[N_1, N_2]$ are symmetric noncooperative equilibria. For very large c, $N_2 < N_1$ so that there is no equilibrium.[10]

PROPOSITION 1: *Symmetric noncooperative equilibrium market structures exist for disutility cost per unit of distance c that is not prohibitively high. Typically there are many such equilibria characterized by the number of traders per market N, which lies in the interval* $[N_1, N_2]$.

We now consider social welfare in this economy. Let the planner's problem be to set up markets so as to maximize the sum of the expected utility of all traders reduced by the transportation costs they incur. In a market of N participants the sum of the expected utilities of all participants minus transportation costs is

$$S(N)=NU(N) -2c(1/2+3/2+\cdots+(N-1)/2).$$

---

[9] Using the definition of $U(N)$ in equation (2), N_1 can be calculated as $N_1=[-\theta(1-\theta)W''(\theta)/2c]^{1/2}$.

[10] Let the left-hand side of (5) be defined as $G(N) \equiv U(N)-c(N-1)/2$. $G(N_1) \gtrless 0 \Leftrightarrow N_1 \gtrless N_2$. Using (2), $G(N_1)=W(\theta)+c(1-3N_1)/2$. Since $N_1>1/3$, the term in parentheses is negative. Therefore, $G(N_1)$ will be negative for large c, implying $N_1>N_2$, and thus there will be no N where both (3) and (5) are satisfied simultaneously.

There are L/N markets of N participants in a line of length L. Thus total surplus is

$$s(N) = [NU(N) - 2c(1/2+3/2 + \cdots + (N-1)/2)]\,L/N$$
$$= [U(N) - cN/4]\,L,$$

proportional to $U(N) - cN/4$, the average utility realized by participants in a market of size N. $s(N)$ is approximately maximized (ignoring integer constraints) at

$$U'(N) = c/4, \tag{6}$$

the solution of which we call N^*.[11] Comparing (4) with (6) and the concavity of $U(N)$ implies $N^* > N_1$. The result that N^* is larger than N_1 is a consequence of liquidity being a positive externality. When c is low, N^* is lower than N_2 and therefore *the surplus-maximizing solution can be achieved as a noncooperative equilibrium.* This is in contrast with the usual result in price-location models as in Nicholas Economides (forthcoming), and Kelvin Lancaster (1979).

Since N^* belongs in the interval $[N_1, N_2]$, there can also exist equilibria with more or fewer traders per market than is efficient. It is striking that the noncooperative equilibrium in this economy may have more traders per market and fewer markets than the social optimum. This possibility arises because liquidity is self-reinforcing. Given an existing equilibrium, new markets may find it impossible to open because nobody wants to use a new market with low liquidity. There may be too few markets to be efficient, $N_2 > N > N^*$, and yet new markets do not open. So even though liquidity is a positive externality, too much liquidity can result from noncooperative behavior!

When c is large, the social optimum lies beyond N_2 traders per market. To achieve the optimal market structure, traders have to be subsidized to participate in larger and fewer markets. This is the case considered by Diamond (1982).

PROPOSITION 2: *The surplus-maximizing market structure is a noncooperative equilibrium when c is low. Noncooperative equilibrium market size can be larger or smaller than is optimal.*

We now briefly consider the effects of addition of traders to the economy. At first sight it may seem that the addition of traders (say through replication) should decrease the variance of price in every market, increase liquidity, and result in higher expected utility for all traders. In fact, the addition of agents may not reduce the variance of price if the uncertainty is location-specific, so that agents at the same location get identical draws. We show in Appendix B that replicating the number of agents when uncertainty is location-specific leaves liquidity per market constant. Thus, traders' utility is also unaffected. However, replication of the number of agents when uncertainty is not location-specific results in increased liquidity *ceteris paribus*. The resulting equilibrium will have denser markets with increased expected utility for all agents.

III. Competitive Equilibrium with Costly Market Services

In many situations market services are not free. We now allow explicit competition in the provision of market services. Instead of having market services provided free of charge, let each market be operated by a real auctioneer (market maker).[12] Suppose that agents can choose between being market makers or traders. When making this choice they are ignorant of the location they will receive on the line if they decide to be traders. After the choice of occupation, the market makers choose their positions on the line to set up their markets. There is no cost to setting up a market except for the opportunity cost of being a trader. The market maker charges each trader a fee F for using his market.[13] After the markets are set up, traders learn their position on the line. Assume that traders are restricted to choose from the set of markets that are offered. Then the traders' decision problem is similar to the one we have discussed before except for the possible difference in fees across markets.

---

[11] Using the definition of $U(N)$ in equation (2), it is straightforward to show that $N^* = [-2\theta(1-\theta)W''(\theta)/c]^{1/2} = 2N_1$. The fact that we use an infinite L in no way affects the validity of the maximization.

[12] Siow (1982) and Townsend (1983) also studied similar costly financial intermediation.

Consider the problem of choice of a fee for the market maker located at m_1. He can lower the fee and hope to attract traders from his competitors, assuming that his competitors will not respond to his price cutting. When he gets additional customers just due to the price cut, his closest competitor to his right, m_2, will lose customers to him. But m_2 will also lose customers to the next market to the right, m_3, because m_2's market is now less liquid. This in turn will drive even more customers away from m_2 toward m_1. Therefore, the number of new customers that m_1 gets from cutting his price depends on the difference in fees and on how liquidity is affected in all other markets. Formally, the new distribution of customers across markets due to m_1's cutting his fee is described by the solution to a second-order difference equation.[14] Market maker m_1 will choose the price that will maximize his profits. But his competitors are doing the same thing, resulting in a Nash equilibrium in fees. Consider the case of symmetrically spaced markets, so that all markets charge the same fee at equilibrium attracting N traders each. Let the equilibrium fee with N traders per market be $F^*(N)$. The wage of a market maker is then $N \cdot F^*(N)$. Since any agent can choose to be a market maker or trader, the wage of the market maker must be equal to the expected utility of a trader from participating in a market of size N. This closes the model and determines the unique market structure in our economy.[15]

We first analyze the game in fees among market makers. After the equilibria of this game are computed, we will return to the (earlier) stage of occupation choice. Formally, let market maker j (operating market m_j) charge a fee F_j for the participation of a trader in his market. For a symmetric positioning of the markets N distance apart, with N traders between every adjacent market, we seek a symmetric noncooperative equilibrium in fees $F^*(N)$.

We now calculate the demand facing market maker $j+1$. Let the marginal trader who is indifferent between going to market m_{j+1} and m_{j+2} be located at $(N_{j+1}-1/2)$ to the right of m_{j+1} (and this implies that he will be $(N-N_{j+1}+1/2)$ away from m_{j+2}). If he goes to market m_{j+1}, there will be $N-N_j+N_{j+1}$ traders in that market and he has to travel $(N_{j+1}-1/2)$ and pay fee F_{j+1}. Therefore, his utility will be $U(N-N_j+N_{j+1})-(N_{j+1}-1/2)c-F_{j+1}$. Similarly, if he goes to market m_{j+2}, there will be $N-N_{j+1}+N_{j+2}+1$ traders at m_{j+2} and his utility will be $U(N-N_{j+1}+N_{j+2}+1)-(N-N_{j+1}+1/2)c-F_{j+2}$. As the trader is on the margin, the following equation has to be satisfied[16]

$$\begin{aligned} U(N-N_j+N_{j+1})-(N_{j+1}-1/2)c-F_{j+1} \\ = U(N-N_{j+1}+N_{j+2}+1) \\ -(N-N_{j+1}+1/2)c-F_{j+2}. \end{aligned}$$

The system of these equations (j integer) determines the marginal consumers and the demand faced by all market makers as functions of the fees charged.

---

[13] The fee is in utility units.

[14] We are grateful to Mike Woodford for his help with the difference equations.

[15] Because the number of agents is finite, the demand and profit functions are discontinuous. We thus establish ε equilibria, where market makers optimize up to ε.

[16] This equation approximates the position of the marginal consumer when the N's are treated as integers. Because we are establishing an ε equilibrium, we can treat this relation (and equations (7)–(9) below) as equalities rather than inequalities.

Let all market makers charge the same fee $F_j = F$ except for one market, say F_1, and let $\Delta F = F_1 - F$. Then the positions of the marginal consumers are the solution of the system of equations:

$$U(N - N_j + N_{j+1}) - (N_{j+1} - 1)c = U(N - N_{j+1} + N_{j+2} + 1) - (N - N_{j+1})c, \qquad j \neq 0, -1, \tag{7}$$

$$U(N - N_0 + N_1) - (N_1 - 1)c - \Delta F = U(N - N_1 + N_2 + 1) - (N - N_1)c, \qquad j = 0, \tag{8}$$

$$U(N - N_{-1} + N_0) - (N_0 - 1)c + \Delta F = U(N - N_0 + N_1 + 1) - (N - N_0)c, \qquad j = -1. \tag{9}$$

Defining $\Delta N_j = N_j - N/2$ and linearizing $U(\cdot)$ around $U(N)$ results in

$$\Delta N_{j+2} + \gamma \Delta N_{j+1} + \Delta N_j = \gamma/2, \qquad j \neq 0, -1, \tag{10}$$

$$\Delta N_2 + \gamma \Delta N_1 + \Delta N_0 = \gamma/2 - \Delta F/U', \qquad (j = 0), \tag{11}$$

$$\Delta N_1 + \gamma \Delta N_0 + \Delta N_{-1} = \gamma/2 + \Delta F/U', \qquad (j = -1), \tag{12}$$

where $\gamma = 2(c - U'(N))/U'(N)$.

In Appendix C we solve the system of (10)–(12). Imposing the condition that perturbations at market 1 should have minimal effect at markets far away from m_1, we show that the general solution of (10) with a convergent path is

$$\Delta N_j = \gamma/2(2+\gamma) + A\rho_1^j, \qquad j = 1, 2, \ldots, \tag{13a}$$

$$\Delta N_j = \gamma/2(2+\gamma) + B\rho_2^j, \qquad j = 0, -1, \ldots, \tag{13b}$$

where $\rho_2 = \rho_1^{-1}$ are the roots of the characteristic equation of equation (10) and $\rho_2 < -1 < \rho_1 < 0$ for $U'(N) < c/2$.[17]

Imposing (11) and (12) determines A and B (see Appendix C) as $A = \Delta F/[\rho_1(1 - \rho_1 - \gamma)U']$ and $B = -\Delta F/[(1 - \rho_1 - \gamma)U']$. Thus, demand for firm 1 when it deviates $\Delta F = F_1 - F$ from the fees of all others is

$$N_1 - N_0 + N = \Delta N_1 - \Delta N_0 + N = N + 2\Delta F/[(1 - \rho_1 - \gamma)U'].$$

The profit function of firm 1,

$$\Pi_1(F_1) = NF_1 + 2(F_1 - F)F_1/[(1 - \rho_1 - \gamma)U'],$$

is concave in F_1, and is maximized at the solution of

$$4F_1/[(1 - \rho_1 - \gamma)U'] - 2F/[(1 - \rho_1 - \gamma)U'] + N = 0.$$

At the symmetric equilibrium $F_1 = F$, which implies that the equilibrium fee is[18]

$$F^*(N) = N(\rho_1 + \gamma - 1)U'/2. \tag{14}$$

Symmetric equilibrium profits of a market

---

[17] We could also have considered agent $N_{j+1} + 1$ (at distance $(N_{j+1} + 1/2)$ from m_{j+1}) being indifferent between going to market m_{j+1} or m_{j+2}. Then the resulting system of equations is similar to (10)–(12) except that $\gamma/2$ is replaced by $-\gamma/2$. As seen next in the text, this change only shifts equally the boundaries of the market for any market maker. Therefore, the demand for market maker $j+1$, being

$$N_{j+1} - N_j + N = \Delta N_{j+1} - \Delta N_j + N,$$

remains unchanged and there will be no effect on the equilibrium.

[18] This analysis was done under the assumption that all consumers participate in the market, that is, that $N(\gamma + \rho_1 - 1)U'/2 = F^*(N) \leq U(N) - c(N - 1)/2$. $F^*(N)$ is positive since it is proportional and of the same sign as $\rho_1 + \gamma - 1 = [c - 2U' + c^{1/2}(c - 2U')^{1/2}]/U' > 0$ because $U' < c/2$.

maker when markets are N apart are

$$(15)\quad \Pi_E(N) = N^2(\rho_1 + \gamma - 1)U'/2$$
$$= N^2\left(c - 2U' + [c^2 - 2cU']^{1/2}\right)/2.$$

This is an increasing function of N.[19]

PROPOSITION 3: *The symmetric fee structure $F_i = F^*(N)$ given by* (14) *is a noncooperative equilibrium of the game among market makers when markets are set N apart. Equilibrium profits are given by* (15).

Now we return to the choice of occupations. The overall equilibrium market structure is determined by the condition that equalizes the expected profits of a market maker with the expected utility of a trader. The expected utility of a trader when he does not know his position on the line (following the analysis of Section III) is $S(N) = U(N) - cN/4 - F^*(N)$, a concave function of N passing through the origin. The equilibrium market structure is determined by the intersection of $\Pi_E(N)$ and $S(N)$. Since $\Pi_E(N)$ is defined for $N > N_0$, where $U'(N_0) = c/2$, the equilibrium N_E exists if $S(N_0) - \Pi_E(N_0) \geq 0$. Equilibrium fees can be calculated by substitution of N_E in (14).

In general, the equilibrium N_E is not the same as the optimal market structure N^*, which is analogous to the one of Section II adjusted for the the market makers' wages. In fact, results for the Cobb-Douglas utility function show that N_E can be smaller or larger than N^*. Thus, even when market makers are aware of the gains from liquidity, and there is competition in fees between market makers and free entry of market makers, the externality caused by liquidity is still not completely internalized. As noted in our introductory discussion, this result is counter to Knight's (1924), where he argued that profit-maximizing ownership of a congested facility leads to efficient pricing.

PROPOSITION 4: *There exists a unique symmetric equilibrium market structure with N_E traders per market, where the ex ante expected utility of an agent is the same in either occupation. Liquidity in this equilibrium may exceed or fall short of the level needed for surplus maximization.*

Note that function $S(N)$ is decreasing in c while $\Pi_E(N)$ is increasing in c.[20] Therefore, the intersection N_E of $S(N)$ and $\Pi_E(N)$ decreases in c. A decrease in the disutility of travel results in fewer and larger markets.

IV. The Monopolist's Solution

So far this paper has only considered competitive market structures. However, many financial markets in the United States, where liquidity considerations are important, are organized by a few financial exchanges. A study of competition between financial exchanges using oligopolistic models of product differentiation is beyond the scope of this paper. As a benchmark, we discuss the market configuration chosen by a monopolist exchange that acts as a market maker in all markets.

The monopolist's problem is particularly relevant for a futures exchange that has to determine the maturity dates of futures contracts in a commodity. Contracts of different maturity dates compete among themselves for liquidity. The extent that a nondiscriminating monopolist can appropriate the market-generated surplus depends on the distance between markets and the level of liquidity in each market. Thus it is a nontrivial choice problem for the monopolist to determine the optimal number of maturity dates and the corresponding fees.

Within the context of our model, the monopolist's objective is to maximize total revenue collected from all markets. Once the monopolist announces the locations of markets and the fee structure, the traders decide whether to participate in a market and in which market to do so. The monopolist will serve all agents on the line. This is because, for any fee structure that leaves some agents at home, the monopolist can bring the markets closer together, close the gap, and increase revenue by establishing a market in the freed space. Thus, the monopolist wants to make the marginal agent, at distance $c(N-1)/2$, indifferent between participating and staying home. From each market the monopolist collects $F(N)\cdot N$. Since the frequency of the markets is $1/N$, total revenues are proportional to $F(N)$. Thus, the monopolist's problem is to

---

[19] For convexity it is sufficient that $U''' < 0$.

[20] $d\Pi_E(N)/c = 1 - 2U' + (c - U')/[c^2 - 2cU']^{1/2} > 0$ for $0 \leq c \leq 1$ because the second term is always positive since $U' < c/2 < c$, and the first term is positive if $U' < c/2 \leq 1/2 \leq c \leq 1$.

$$\underset{N}{\text{Maximize}}\ F(N)$$

subject to

$$U(N)-c(N-1)/2-F(N)\geq 0,$$

that is, that traders come to the market rather than stay home, and also subject to

$$(4)\qquad U'(N)\leq c,$$

that is, that the symmetric equilibrium fee structure of the monopolist is a noncooperative equilibrium for traders. It is equivalent to

$$(16)\qquad \underset{N}{\text{Maximize}}\ U(N)-c(N-1)/2$$

$$\text{subject to } U'(N)\leq c.$$

Its solution is at N_M defined by

$$(17)\qquad U'(N)=c/2.$$

Clearly N_M is in (N_1,N_2). In comparison with the surplus-maximizing outcome, the monopolist will operate a larger number of smaller markets.[21] This is because the surplus-maximizing outcome is defined by $U'(N^*)=c/4$ (equation (6)) and $U(\cdot)$ is concave.

Comparing the monopolist's market structure with the equilibrium of independent market makers of Section IV, we see that $U'(N_M)=c/2>U'(N_E)$. By the concavity of $U(\cdot)$, $N_M<N_E$. The monopolist will open smaller and more numerous markets than independent market makers.

PROPOSITION 5: *A monopolist will operate smaller and more numerous markets than independent competing market makers. Further, his markets are always smaller and more numerous than is optimal.*

Lacking the ability to price discriminate and appropriate the whole surplus, the monopolist avoids creating large markets with high surplus. Instead he institutes a large number of smaller markets where he can appropriate a larger percentage of the surplus. We note that *the overcrowding of the space with markets happens despite the fact that there is no threat of entry.* Overcrowding of the product space to deter potential entrants has been noted by Richard Schmalensee (1978), among others.

V. Discussion

Consider replacing each spot market in this economy with state-contingent claims markets (for example, Debreu, 1959, ch. 7) for agents that go to the same market location. Agents will still go to a specific market location. But each agent now trades in state-contingent claims with other agents at the same location before he knows the realization of endowments. Every agent at the same market location is *ex ante* identical. So every agent will have the same excess demand functions for these contingent claims commodities. The equilibrium prices must be such that all agents will have the same *ex post* consumption bundle. Therefore, when k is realized, each agent at the same market will consume $(1-k,k)$. Assuming that the representative agent has a utility function which is concave in the two goods, his indirect utility function will also be concave in

---

[21] Using the definition of $U(N)$ in equation (2), it is easy to show that $N^*/N_M=\sqrt{2}$ so that the monopolist operates approximately 40 percent more markets than is optimal.

k.[22] Substituting this indirect utility function in equation (2) in the place of $W(k)$, one can derive an expected benefit function with the same properties as $U(N)$. The results of the rest of the paper follow. Therefore, the qualitative features of the symmetric noncooperative equilibria are the same whether agents are faced with spot markets or state-contingent claims markets.

The difference between the state-contingent claims markets in this economy and in the standard general equilibrium model is that agents in this economy must go to a specific market location before they can participate in the state-contingent claims markets. There is risk sharing within a market location but not across market locations. In the standard model, agents can participate in a complete market structure without first having to go to any specific location. The standard model allows risk sharing among all individuals in the economy, whereas our model only allows risk sharing among endogenous subsets of individuals.

The alternative market setup considered above shows that it is not the spot market setup of our problem that is important. The normative results on liquidity is due to the Nash equilibrium concept that we employ. Some readers have questioned whether alternative equilibrium concepts, such as the core (as used in studies on financial intermediation by John Boyd and Edward Prescott, 1986, and Townsend, 1983), may give different normative results. In our problem, the core is efficient because it allows a redistribution of income from agents close to a market to those who are far away. The redistribution means that all agents who go to the same market will have the same *ex ante* utility, making them indifferent to their distance from the market. The efficient market structure allows for the largest surplus to be redistributed, which means no other coalition can be formed that will satisfy all agents in this other coalition. Therefore, the core will provide the efficient level of liquidity. The main difficulty with using the concept of the core in our problem is that the redistribution that is necessary seems difficult to enact for many relevant problems. In particular, the identity of every agent (that is, his location) has to be common knowledge to all agents in order to implement the core allocation. The identity of agents is not necessary for constructing the Nash equilibria. If the identities of agents were somehow known, it is possible that price discrimination in fees by market makers within the Nash equilibrium construct might mimic the core solution.

An area in which the model with free market services might apply is in the choice of standards when there is a variety of products. For example, the personal computer industry has a large variety of potential and actual products. In this industry a few standards have already arisen. Many consumers buy an IBM or IBM-compatible machine, even though it is not the "best" for what they currently want to do. Other products may be able to do what they want better and at a lower price. However, most consumers know that they may use the computer to solve other problems in the future. By buying "the standard," they are buying insurance that accessories (software and hardware) will be available for solving those problems. On the other side of the market, firms may not produce products which accomplish a task most efficiently, but will rather produce IBM-compatible products. Since firms do not have to pay a fee in choosing the "standard," there will be a range of indeterminancy for the equilibrium standards, as predicted by our model. The standards that obtain in an actual industry can often be predicted by participants in that industry from their knowledge of initial conditions. For example, most informed observers expected IBM to become a standard in the personal computer industry. This observation is not inconsistent with the fact that at some point "the standard" may look arbitrary, given the available knowledge and technology in that society. Research on network externalities considers closely related issues (for example, Dennis Carlton and Mark Klamer, 1983; Joseph Farrell and Garth Saloner, 1985; Michael Katz and Carl Shapiro, 1985). Recent work (Economides, 1987; Carmen Matutes and Pierre Regibeau, 1986) has also demonstrated that firms may have incentives to produce compatible products, even in the absence of network externalities.

---

[22] In this setup $W(k)=w(1-k,k)$ where $w(\ \ ,\ \cdot)$ is concave. Then under regularity $W''(k)=w_{11}+w_{22}-2w_{12}$. For concavity of $W(\cdot)$ we need to show $|w_{11}|+|w_{22}|+2w_{12}>0$. This is obviously true for $w_{12}>0$. For $w_{12}<0$, it is sufficient to show that $|w_{11}|+|w_{22}|-2|w_{12}|>0$. By concavity of w, we know that $(|w_{11}w_{22}|)^{1/2}>|w_{12}|$. Thus $|w_{11}|+|w_{22}|-w|w_{12}|>|w_{11}|+|w_{22}|-2(|w_{11}w_{22}|)^{1/2}=(|w_{11}|^{1/2}-|w_{22}|^{1/2})^2\geq 0$.

Another example where free standards matter is the agents' hours of work in an economy. Workers have to interact with other workers and clients both within and outside the firm. Often these interactions are not mutually anticipated or coordinated. Our model suggests that workers with heterogeneous preferences will tend to work the same schedule. Moreover, the comparative statics results in Appendix B suggest that there will be more numerous and diverse types of work schedules in a large city, where there is a larger number of types of workers and more workers of each type. Empirical evidence in Siow (1987) shows that a worker is penalized for not coordinating his hours of work with his co-workers.

APPENDIX A

A solution of the problem when agents have Cobb-Douglas utility functions follows. Let P be the relative price of good y with respect to x. A consumer with utility function $U(x, y) = x^{\alpha}y^{\beta}$, when endowed with A units of x, has budget constraint $x + Px = A$ and (gross) demands $x_1 = \alpha A/(\alpha+\beta)$, $y_1 = \beta A/(\alpha+\beta)P$. When endowed with A units of y he has budget constraint $x + Py = AP$ and demands $x_2 = \alpha AP/(\alpha+\beta)$, $y_2 = \beta A/(\alpha+\beta)$. Market clearing implies $x_1 X + x_2(N - X) = XA \Leftrightarrow P = \beta X/\alpha(N - X) = \beta(1-k)/\alpha k$. The equilibrium indirect utility function of a consumer endowed with A units of x participating in a market (N, k) is $V_1(k) = (\alpha A/(\alpha+\beta))^{\alpha+\beta}(k/(1-k))^{\beta}$. For a consumer endowed with A units of y, the corresponding indirect utility is $V_2(k) = (\beta A/(\alpha+\beta))^{\alpha+\beta}((1-k)/k)^{\alpha}$. Let $W(k) = (1-k)V_1(k) + kV_2(k)$ as in equation (1) in the text. A straightforward calculation will show that $W(k)$ is concave as long as α, β are in the open interval $(0,1)$. When the utility functions are CES, $U(x, y) = (x^{\rho} + y^{\rho})^{1/\rho}$, $\rho < 1$, a type 1 consumer, facing constraint, $x + Py = A$, will demand $x_1 = A/(1+P^r)$, $y_1 = AP^{r-1}/(1+P^r)$, where $r = \rho/(\rho-1)$. Similarly a type 2 consumer, facing constraint $x + Py = AP$, will demand $x_2 = AP/(1+P^r)$, $y_2 = AP^r/(1+P^r)$. Market clearing implies $x_1 X + x_2(N - X) = XA \Leftrightarrow$

$$\text{(A1)} \qquad 1 - k + kP = (1-k)(1+P^r)$$

$$\Leftrightarrow P = (k/(1-k))^{1/(r-1)}.$$

The indirect utility functions are $V_1 = A(1+P^r)^{-1/r}$, $V_2 = AP(1+P^r)^{-1/r}$. Thus, $W(k) \equiv (1-k)V_1 + kV_2 = (1-k+kP)V_1 = (1-k)A(1+P^r)^{1-1/r}$, using (A1). Substituting the clearing price we have

$$W(k) = A(1-k)\left[1+(k/(1-k))^{r/(r-1)}\right]^{(r-1)/r}$$

Direct computation reveals

$$W'(k) = A\left[1+(k/(1-k))^{r/(r-1)}\right]^{-1/r} \times\left[-1+(k/(1-k))^{1/(r-1)}\right],$$

and

$$W''(k) = A(k/(1-k))^{1/(r-1)} \times\left[1+(k/(1-k))^{r/(r-1)}\right]^{-1-1/r} / \left[(r-1)k(1-k)^2\right] < 0,$$

where all terms except the denominator are positive. $r-1$ is negative for all $\rho < 1$, that is, for the whole range of definition of the CES.

APPENDIX B

Suppose we double all agents at the old positions. First, let us consider the case when uncertainty is not location-specific. Let any two agents at the same location get independent draws from the distribution of endowments. At every market, the distribution of types is preserved. Given k, the equilibrium price is unaffected and so are the indirect utility functions $V_1(k), V_2(k)$ of each trader and their weighted sum $W(k)$. The *function* $U(N)$ is the same as before, but now it has to be evaluated at $2N$. Let markets be $\hat{N}_1 d$ apart, at the lower bound of the equilibrium existence region (where the corresponding number before doubling was $N_1 d$). There are now $2\hat{N}_1$ traders per market. Equation (4) for this equilibrium is

$$\text{(A2)} \qquad U'(2\hat{N}_1) = cd,$$

which implies $2\hat{N}_1 = N_1$, or $\hat{N}_1 = N_1/2$, since N_1 solves $U'(N_1) = cd$. Thus, when agents are doubled at the old locations and receive independent draws, the number of agents per market remains unaffected ($2\hat{N}_1 = N_1$), while the markets are twice as dense at distances $\hat{N}_1 = N_1 d/2$ (compared with $N_1 d$ originally). The expected utility of agents is now higher because they have to travel half the distance to find a market of the same liquidity as before doubling. Alternatively, suppose that all uncertainty is location-specific, such as uncertainty associated with the local weather. After replication, let agents receive identical endowments. Again $V_1(k)$, $V_2(k)$, and $W(k)$ are unaffected. Further, the variance of k is the same as before replication, although there are now twice as many agents per market. The original expected utility function

$$\text{(A3)} \quad U(N) = W(\theta) + W''(\theta)\theta(1-\theta)/(2N)$$

has been replaced by a new utility function for the replicated model

$$U_1(2N) = W(\theta) + W''(\theta)\theta(1-\theta)/(2N),$$

and therefore

$$U_1(N) \equiv U(N/2).$$

Let markets be $\hat{N}_1'd$ apart at the lower bound of the equilibrium existence region. There are now $2\hat{N}_1'$ traders per market. Equation (4) is now

$$U_1'(2\hat{N}_1') = cd \Leftrightarrow U'(\hat{N}_1') = cd \Leftrightarrow \hat{N}_1' = N_1.$$

The resulting markets are at the same distance as before replication, with twice as many traders per market, but all traders receive the same utility as before, $Z_1(2N, \alpha) = U_1(2N) - c\alpha = U(N) - c\alpha = Z(N, \alpha)$.

We have shown that replicating the number of agents when uncertainty is location-specific leaves liquidity per market constant. Thus, traders' utility is also unaffected. However, replication of the number of agents when uncertainty is not location-specific results in increased liquidity *ceteris paribus*. The resulting equilibrium will have denser markets with increased expected utility for all agents.

APPENDIX C

The homogeneous version of equation (10), $\Delta N_{j+2} + \gamma \Delta N_{j+1} + \Delta N_j = 0$, has the characteristic equation $\rho^2 + \gamma\rho + 1 = 0$ with solutions $\rho_1 = (-\gamma + (\gamma^2 - 4)^{1/2})/2$, $\rho_2 = 1/\rho_1$, or equivalently, since $\gamma = 2(c - U')/U'$, $\rho_1 = (U' - c + [c^2 - 2cU']^{1/2})/U'$, $\rho_2 = (U' - c - [c^2 - 2cU']^{1/2})/U'$. The roots are real distinct for $U' < c/2$. For $U' = c/2$, they are real and coinciding $\rho_1 = \rho_2 = -1$. Then the solution of the homogeneous equation $\Delta N_j = A(-1)^j + Bj(-1)^j$ diverges as j goes to infinity. For $U' > c/2$, the roots are complex and the solution is $\Delta N_j = A\cos(\theta j) + B\sin(\theta j)$, where $\cos\theta = -\gamma/2 = (U' - c)/U'$, an exact oscillation. In the last two cases a disturbance at $j = 1$ has large effects at markets far away, an event we rule out. In the case of distinct roots arising from $U' < c/2$, it is easy to see that $\rho_2 < -1 < \rho_1 < 0$. The general solution of the homogeneous equation for markets to the right of m_1, $\Delta N_j = A_1\rho_1^j + P_2\rho_2^j$, converges as $j \to \infty$, if and only if A_2 is zero. Similarly, the general solution for markets to the left of m_1, $\Delta N_j = B_1\rho_1^j + B_2\rho_2^j$, converges as $j \to -\infty$, if and only if B_1 is zero. Hence the convergent solution is

$$\Delta N_j = A\rho_1^j, \qquad j = 1, 2, \ldots, \text{ and}$$

$$\Delta N_j = B\rho_2^j, \qquad j = 0, -1, \ldots.$$

The inhomogeneous equation (10) has a particular solution $\Delta N_j = \gamma/2(2+\gamma)$.

Thus the solution of (10) is

$$\Delta N_j = \gamma/2(2+\gamma) + A\rho_1^j, \qquad j = 1, 2, \ldots,$$

$$\Delta N_j = \gamma/2(2+\gamma) + B\rho_2^j, \qquad j = 0, -1, \ldots.$$

Conditions (11) and (12) can thus be written as

$$A\rho_1^2 + A\gamma\rho_1 + B = -\Delta F/U'$$

$$A\rho_1 + B\gamma + B\rho_1 = \Delta F/U',$$

which are solved by $A = \Delta F/[(1 - \rho_1 - \gamma)U'$, $B = -\Delta F/[(1 - \rho_1 - \gamma)U']$, so that the solution of the system is

$$\Delta N_j = \gamma/2(2+\gamma) + \Delta F\rho_1^{j-1}/[(1 - \rho_1 - \gamma)U'],$$
$$j = 1, 2, \ldots$$

and

$$\Delta N_j = \gamma/2(2+\gamma) - \Delta F\rho_2^j/[(1 - \rho_1 - \gamma)U'],$$
$$j = 0, -1, \ldots.$$

REFERENCES

Black, Deborah G., "Success and Failure of Futures Contracts: Theory and Empirical Evidence," unpublished manuscript, New York University, 1985.

Boyd, John and Prescott, Edward, "Financial Intermediary-Coalitions," *Journal of Economic Theory*, April 1986, *38*, 211–32.

Braid, Ralph, "Duopoly Pricing of Congested Facilities," Discussion Paper No. 322, Department of Economics, Columbia University, 1986.

Carlton, Dennis, "Futures Markets: Their Purpose, Their History, Their Growth, Their Success and Failures," *Journal of Futures Markets*, Fall 1984, *4*, 237–72.

______ **and Klamer, Mark,** "The Need for

Coordination among Firms, with Special Reference to Network Industries," *University of Chicago Law Review*, Spring 1983, *50*, 446–65.

Cass, David and Shell, Karl, "Do Sunspots Matter?," *Journal of Political Economy*, April 1983, *91*, 193–227.

Debreu, Gerard, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New York: Wiley & Sons, 1959.

Diamond, Peter, "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, October 1982, *90*, 881–94.

______, *A Search-Equilibrium Approach to the Micro Foundations of Macroeconomics*, Cambridge: MIT Press, 1984.

Economides, Nicholas, "The Principle of Minimum Differentiation Revisited," *European Economic Review*, September 1984, *24*, 345–68.

______, "Equilibrium Existence and Optimality in Differentiated Product Markets," *Journal of Economic Theory*, forthcoming.

______, "Desirability of Compatibility in the Absence of Network Externalities," Discussion Paper No. 351, Department of Economics, Columbia University, 1987.

Farrell, Joseph and Saloner, Garth, "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, Spring 1985, *16*, 70–83.

Garbade, Kenneth and Silber, William, "Structural Organization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk," *Journal of Finance*, June 1979, *34*, 577–93.

Katz, Michael and Shapiro, Carl, "Network Externalities, Competition, and Compatibility," *American Economic Review*, June 1985, *75*, 424–40.

Knight, Frank, "Some Fallacies in the Interpretation of Social Cost," *Quarterly Journal of Economics*, August 1924, *38*, 582–606.

Lancaster, Kelvin, *Variety, Equity, and Efficiency*, New York: Columbia University Press, 1979.

Lippman, Steven and McCall, John, "An Operational Measure of Liquidity," *American Economic Review*, March 1986, *76*, 43–55.

Mortensen, Dale, "Job Matching under Imperfect Information," in Orley Ashenfelter and J. Blum, eds., *Evaluating the Effects of Social Programs*, Princeton, NJ: Princeton University Press, 1976.

Matutes, Carmen and Regibeau, Pierre, "Mix and Match: Product Compatibility without Network Externalities," University of California-Berkeley, 1986.

Schmalensee, Richard, "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry," *Bell Journal of Economics*, Autumn 1978, *9*, 305–27.

Siow, Aloysius, "Extending Diamond's Model of a Search Economy," Discussion Paper No. 143, Department of Economics, Columbia University, New York, NY, 1982.

______, "The Use of Wages in Coordinating Hours of Work," unpublished manuscript, 1987.

Telser, Lester G., "Why There Are Organized Futures Markets," *Journal of Law and Economics*, April 1981, *24*, 1–22.

Townsend, Robert, "Theories of Contract Design and Market Organization: Conceptual Bases for Understanding Futures Markets," *The Industrial Organization of Futures Markets*, in Ronald W. Anderson, ed., Lexington: Lexington Books, 1984.

______, "Theories of Intermediate Structures," in Karl Brunner and Allan-Meltzer, eds., *Carnegie-Rochester Conference Series on Public Policy*, Spring 1983, *18*, 221–72.

On The Strategic Choice of Spatial Price Policy

By JACQUES-FRANÇOIS THISSE AND XAVIER VIVES*

Price discrimination emerges as the unique equilibrium outcome in games with either simultaneous choice of policy and price or sequential choice in which firms may commit first to uniform mill pricing before the actual market stage. Our results are used to analyze some common business practices that arise in geographical pricing, like the basing point system, and in the pricing of varieties or options from a base product in a product differentiation context.

In any industry, current business practices are not only the result of an adaptation by firms to particular technological or institutional conditions but also reflect, or are the instruments of, a *strategic* positioning of firms in the market.[1] The specific price policies that firms follow are obviously one of the most important business practices in any market environment. Although their strategic importance is widely recognized in the business literature, perhaps somewhat surprisingly, the economics literature in general has not paid enough attention to the detail of specific pricing methods and, consequently, has not been able to provide convincing explanations of the incentive structure that lies behind certain price policies. In this paper, we would like to present some very simple game-theoretic models which are naturally related to some pricing policies found in geographical or product differentiation contexts. Let us begin by describing some of these policies.

In a geographical context, examples of alternative pricing methods are: (*i*) the zone price system, under which a specific delivered price is charged to all buyers located in a given region, such as for plasterboard in the United Kingdom or cement in Belgium; (*ii*) the (single) basing point system, in which the delivered price equals a base price plus the cost of shipping to the place of delivery from a given basing point that need not be where the seller is actually located, such as the Pittsburgh-plus system used in the steel industry in the United States until the mid-1920s or the Portland-plus system used for plywood in the United States until the 1970s; (*iii*) uniform free on board (FOB) prices, in which the delivered price equals the mill price—the same for all customers—plus the actual transport costs.[2]

In a product differentiation context, the choice of a price policy is, in a certain sense, closely related to the firm's variety offer. For example, car manufacturers may decide to provide a single standardized product to satisfy a hypothetical average consumer, as shown by the historical examples of the Ford T and the Volkswagen's "Beetle." Alternatively, they may offer a basic product with a series of options with different price tags such as most car manufacturers currently stock.[3] The first case corresponds to a uniform (FOB) price policy, in which the manufacturers leave the options to indepen-

---

*CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, and University of Pennsylvania, Philadelphia, PA 19104, respectively.

[1] We are grateful to Jonathan Hamilton, Louis Phlips, Kriss Sjoblom, and an anonymous referee for helpful comments. The authors' research was supported respectively by the National Science Foundation grant no. SES-8502886 to the University of Pennsylvania and by the Research Fund, University of Pennsylvania.

[2] See F. M. Scherer, 1980, ch. 11; L. Phlips (1983), ch. 1, and M. L. Greenhut et al., 1987, for a detailed description of spatial pricing policies. Actual pricing patterns are studied in Greenhut et al., 1980, and Greenhut, 1981.

[3] The number of options may be very large indeed. The different makes of the Ford Escort, bumper-to-bumper, would cover the distance between Paris and Brussels, that is, 230 km (see L. Augier and J.-P. Icikovics, 1982).

dent producers.[4] The second case, in which the options are under the manufacturers' control, typically involves discriminatory pricing.

A basic question, which interests both the analyst and the policymaker, is to determine what kind of price policy the government should encourage. In particular, is price discrimination necessarily evidence of lack of competition and harmful to the consumers?

In a geographical context, there is price discrimination when firms do not set uniform FOB prices, that is, when the difference between delivered prices at two distinct locations does not equal the transport cost between them. In a product differentiation context, we say that there is price discrimination if two varieties are sold at different base prices, that is, prices corrected by the cost of the corresponding options. Many countries have legislated against price discrimination defined as charging different prices to different buyers of the same goods. Obviously, the problem is to know whether the goods sold are the same since a given physical good is a different economic good depending on location, date of delivery, or service features, for example. In the United States, the Robinson-Patman Act prohibits charging different prices to different buyers of "goods of like grade and quality" where the result "may be substantially to lessen competition or then to create a monopoly in any line of commerce, or to injure, destroy, or prevent competition with any person who either grants or knowingly receives the benefit of such discrimination, or with customers of either of them" (see Scherer, 1980, p. 572). In the United Kingdom, the now-abolished Price Commission has favored uniform FOB pricing as shown, for example, by the series of reports published in 1978. In any case, it has been claimed that the motivation for these laws was more to limit than to promote competition (see, for example, Scherer, 1980, ch. 21) and that they may have perverse effects (see, for example, M. Schwartz, 1986).

The importance of price policies for the firms and the public authority is clear. Nevertheless, there exist very few models which rationalize in economic terms specific price policies, like the basing point system for example, or which explain the strategic incentives that firms may have to price discriminate. A great deal of attention has been paid to the issue of uniform versus discriminatory pricing in the general literature on *monopolistic price discrimination* (see, for example, W. Y. Oi, 1971; A. C. Pigou, 1920; J. Robinson, 1933; R. Schmalensee, 1981) and in the theory of *spatial monopoly* (see M. J. Beckmann, 1976; Greenhut and H. Ohta, 1972; W. L. Holahan, 1975, for example). The general conclusion is that a monopolist can secure higher profits by price discriminating. On the other hand, there is no such clear-cut statement for the consumers. Very often (and this is true in the spatial monopoly case), they are partitioned under discriminatory pricing into those who pay a higher price and those who pay a lower price. However, as a whole, consumers often prefer uniform FOB pricing.

In contrast, much less attention has been devoted to the choice of a price policy in an oligopolistic environment in which strategic considerations are of primary importance.[5] The purpose of this paper is to contribute to the analysis of this issue. More specifically, we want to examine (*i*) the incentives that arise in spatial competition for firms to price discriminate or price uniformly, and (*ii*) the consequences of the price policy choice for firms and consumers.

In our paper, we consider a duopoly in which each firm is already located in some (geographical or characteristics) space and has to choose whether to price discriminate or not, taking the strategic effects of its choice into account. We deal, first, with the simultaneous choice of policy and price (Section II) and, second, with the case where firms may commit to a price policy first and then compete in prices contingent on the chosen policies (Section III).

---

[4] Alfred P. Sloan, Jr., in the chapter about styling in his book "*My Years with General Motors*" talks about the custom body shop of Don Lee of Los Angeles, CA, "in which he built special bodies on both foreign and American chassis for Hollywood movie stars and wealthy people of California" (A. Sloan, 1972, p. 312).

[5] Most contributions in spatial competition take the pricing policy as given. A notable exception is J. Greenhut and Greenhut (1975). However, these authors study spatial discrimination with quantity-setting firms. Here, we consider price-setting firms. A comparison of these two forms of spatial discrimination is contained in J. Hamilton et al. (1987).

When firms do not commit to a price policy, they compete with unrestricted price schedules, choosing a (delivered) price for every location. We are interested in knowing whether uniform (FOB) pricing can arise in equilibrium in such a situation. Under some circumstances, firms do commit themselves to a specific price policy. Such is the case, for example, in the basing point system. Suppose that there are two firms established at two different places and one basing point. The firm at the basing point is a price leader and announces a (uniform) base price. The other firm reacts optimally to the leader's choice by matching the full price (base price plus transport cost) wherever possible, that is, by price discriminating. In this situation, the leader prices uniformly and the follower price discriminates. In other words, the leader is committed to uniform FOB pricing. One question we want to answer is then: Can this system be rationalized in a context in which there is no scope for collusion, that is, in a two-stage game where, first, firms may commit to certain pricing policies, and then, compete in prices according to the selected polices?

In both the simultaneous no-commitment and the sequential commitment cases, *price discrimination* (by all firms in the market) *emerges as the unique equilibrium outcome*, even though it may well be that firms would make more profits by following a uniform price policy. This is so because spatial discriminatory pricing gives more flexibility to a firm to respond to its rival's actions. But then firms may get trapped into a Prisoner's Dilemma-type situation and end up with lower profits due to the intense competition unleashed. Contrary to general belief, uniform (FOB) pricing is therefore *not* evidence of a more competitive environment.[6]

With simultaneous choice of policy and price the result is very general in terms of the characteristics space, density of demand, and transportation cost schedules which are allowable. With sequential choice much more restrictive assumptions are made in order to get existence and easy characterization of subgame-perfect equilibria (the assumptions are in any case usual in the literature). Section I presents the model, and our conclusions are given in Section IV.

I. The Model

We consider two firms $i=1,2$ selling a homogeneous product. Firm i is located at point y_i of the n-dimensional space $\mathbb{R}^n$ (when the model is interpreted in the geographical context, we have $n=2$) and produces the product at a constant marginal cost c_i. It is supposed that firms are not located at the same point, that is, $y_1 \neq y_2$. Consumers are continuously distributed over a compact subset X of $\mathbb{R}^n$. The density of demand for the product at $x \in X$ is given by a (measurable) function $f(p,x)$ of the full price p (that is, the price gross of transportation costs) paid by the consumers, and of the location x. The transportation cost of one unit of the product is given by a strictly increasing nonnegative function $t_i(\|y_i - x\|)$ of the distance $\|y_i - x\|$, where $\|\cdot\|$ is a norm defined on $\mathbb{R}^n$, with $t_i(0)=0$. In the geographical context typically, because of scale economies in transportation, t_i is a concave function of distance. Examples of norms that can be used are the Euclidean and the Manhattan norms.[7]

Two price policies are considered: uniform (U) and discriminatory (D). In the geographical context, uniform FOB pricing means that firm i charges the same mill price p_i to the consumers irrespective of their location. In this case, the full price of firm i at $x \in X$ is equal to the mill price plus the transportation cost, that is, $p_i(x) = p_i +$

---

[6] See also G. Norman (1983).

[7] Other examples of norms used in location theory are the l_p-norms (see R. F. Love and J. G. Morris, 1979) and the block norms (see J. E. Ward and R. E. Wendell, 1985).

$t_i(\|y_i - x\|)$. This is so, for instance, because the transportation is under the control of the consumers who use the services of independent carriers charging $t_i(\cdot)$. In the rest of the paper, we refer to uniform FOB pricing as *uniform pricing.*[8] *Discriminatory pricing* occurs when firm i bears the transportation cost and chooses a price schedule $p_i(\cdot)$ which describes the delivered price $p_i(x)$ at which firm i is willing to supply consumers at location $x \in X$. The full price of firm i at x is now given by $p_i(x)$. The mill price effectively paid by the consumers, that is, $p_i(x) - t_i(\|y_i - x\|)$, generally changes with their location x. In other words, the firm discriminates among consumers on the basis of their location.

When the model is interpreted as a model of product differentiation, uniform pricing is equivalent to firm i's selling a *single* product located at y_i (in the characteristic space) and consumers paying a full price consisting of the price p_i of the product plus the cost $t_i(\|y_i - x\|)$ incurred by the consumers in using the services of independent producers who adapt firm i's product to their requirements given by x. Now t_i is better viewed as a convex function of distance. In the discriminatory case, firm i bears the cost $t_i(\|y_i - x\|)$ of redesigning its basic product (y_i) and offers the whole *band* of varieties. As product design is under its control, the firm may discriminate among consumers on the basis of their requirements. We say that price discrimination occurs when the price difference between two varieties does not correspond to the difference in the respective costs of redesigning the basic product.

A slightly modified version of the celebrated Hotelling model is a good example to illustrate the above general model. Let us first consider the geographical interpretation. Two sellers of a homogeneous product are located at the endpoints of Main Street; geometrically, $y_1 = 0$, $y_2 = 1$, and X is the segment $[0,1]$. They have identical and constant marginal production costs. Customers are uniformly distributed along Main Street and have fixed and identical requirements for the product. Finally, transportation costs are linear in distance. If both sellers follow a uniform price policy (as in Hotelling), consumers pay the price at the firm's door where a consumer buys the product plus the transportation cost for delivery. On the other hand, when sellers deliver the product (as the pizzaman), they may discern between consumers and price discriminate with respect to locations by absorbing part of the transportation costs.

Let us now come to the characteristics interpretation. Following Hotelling, we suppose that the segment $[0,1]$ describes the sweetness of cider: $y_1 = 0$ means sour cider and $y_2 = 1$ sweet cider. Under uniform pricing, firms produce only the two extreme products, and consumers adopt these products to their most preferred level of sweetness at a cost corresponding to the transportation cost in the geographical approach (how to change the sweetness of cider is a technical detail that we leave to the imagination of the reader). On the contrary, under discriminatory pricing, the firms offer the whole spectrum of sweetness and price each variety.

II. Simultaneous Choice of Policy and Price

In this section we investigate whether discriminatory or uniform pricing arises in equilibrium when firms choose simultaneously pricing policy and price, that is, when firms compete in price schedules which are unrestricted. A strategy for firm i is then a *price schedule* $p_i(\cdot)$ that specifies the delivered price at which firm i is willing to supply consumers at location x in X. The delivered price at x must cover the total (production plus transport) marginal cost. If firm i were to price *below* total marginal cost it could do at least as well, for any given price of the rival, by pricing *at* marginal cost. Formally, we assume $p_i(\cdot)$ to be in the set

$$\mathscr{P}_i \equiv \{\, p_i(\cdot) \text{ a nonnegative function defined on } X, \text{ measurable and such that, for all } x \in X,\ p_i(x) \geq c_i + t_i(\|y_i - x\|)\}.$$

---

[8] Let us emphasize the fact that what we call here uniform pricing is different from uniform delivered pricing as defined in postage stamp systems.

The *potential market area* of firm i is the set of locations at which the firm faces a positive demand density when pricing at total marginal cost $m_i(x) \equiv c_i + t_i(\|y_i - x\|)$, that is,

$$A_i \equiv \{x \in X; f[m_i(x),x] > 0\}.$$

The most interesting case occurs when the set of consumers who consider buying from either firm, that is, $A_1 \cap A_2$, is nonnegligible (technically a nonzero measure set). Since the product is homogeneous, each consumer purchases from the firm with the lower delivered price. In the event of a price tie, we assume that consumers do the socially optimal thing and buy from the firm with the lower production and transportation cost. This may be rationalized noting that this firm can always price ε below its rival's total marginal cost. If, for some consumer location x, both firms have the same total marginal costs and charge the same delivered price, they split the local demand. Generically, the set of locations for which $m_i(x) = m_j(x)$ is negligible. To ease notation we will assume this to be the case.

Given the strategies $p_i(\cdot)$ and $p_j(\cdot)$ of the two firms, the *market area* of firm i, $M_i(p_i(\cdot), p_j(\cdot))$, is then the set of locations in the potential market area A_i for which, either firm i quotes the lower-delivered price or, if both firms quote the same price, firm i has the lower-total marginal cost. Firm i's *profits* are, therefore, equal to

$$\Pi_i(p_i(\cdot), p_j(\cdot)) = \int_{M_i(p_i(\cdot), p_j(\cdot))} [p_i(x) - m_i(x)] \times f[p_i(x), x]\, dx.$$

We thus have a well-defined game with strategy sets P_i and payoffs Π_i, $i = 1,2$. To analyze the Nash equilibria of this game, we need some additional assumptions and definitions.

Let us assume that, for each location x in X, the demand density $f(p,x)$ is continuous and downward sloping in p, and that $pf(p,x)$ is bounded from above. This implies that there exists a monopoly price for every x in X which, for simplicity, we assume unique. Denote by $p_i^M(x)$ the monopoly price of firm i at location x, that is,

$$p_i^M(x) = \operatorname*{argmax}_{p_i} \{(p_i - m_i(x)) f(p_i, x)\}.$$

We say that firm i has a *monopoly position* at location x whenever its monopoly price at x does not exceed the total marginal cost of the rival firm, that is, $p_i^M(x) \leq m_j(x)$, $i \neq j$.

Suppose now that firm i does *not* have a monopoly position at x. We say that firm i has a *cost advantage* at location x if its total marginal cost is lower than that of its rival, that is, $m_i(x) < m_j(x)$. Let x be such a location and define $\bar{p}_i(x)$ as the profit-maximizing price at x when firm i cannot charge more than $m_j(x)$, that is,

$$\bar{p}_i(x) \in \operatorname*{argmax}_{p_i} \{(p_i - m_i(x)) f(p_i, x); p_i \leq m_j(x)\}.$$

Without significant loss of generality, we may assume that $\bar{p}_i(x)$ is unique. Notice that when the profit $(p_i - m_i(x)) f(p_i, x)$ is quasi-concave in p_i, $\bar{p}_i(x)$ takes its highest possible value, namely $\bar{p}_i(x) = m_j(x)$. Otherwise, the profit is not single-peaked and there may be a local maximum at a price less than $m_j(x)$.

We claim that there exists an equilibrium for the above game in which firm i chooses a price schedule given by

$$p_i^*(x) = \begin{cases} p_i^M(x), & \text{if firm } i \text{ has a monopoly position at } x \\ \bar{p}_i(x), & \text{if firm } i \text{ does not have a monopoly position at } x \text{ but has a cost advantage} \\ m_j(x), & \text{otherwise.} \end{cases}$$

The argument is as follows. Since the marginal production cost of firm i is constant and since the transportation cost to a point is unaffected by transportation to other points, it is sufficient to show that, given $p_j^*(x)$, firm i maximizes its profit density at (almost) every x in X by setting $p_j^*(x)$. We have:

(i) If firm i enjoys a monopoly position at x, it will serve the local market and will reach its highest possible profits by monopoly pricing since $m_i(x) \leq p_i^M(x) \leq m_j(x)$.

(ii) If firm i does not have a monopoly position but has a cost advantage at x, then by pricing at $\bar{p}_i(x)$ it captures the local demand, since $\bar{p}_i(x) \leq m_j(x)$ and $m_i(x) < m_j(x)$, (recall that according to our convention if the two firms quote the same price consumers buy from the firm with lower total marginal cost) and maximizes its profits at x by definition of $\bar{p}_i(x)$.

(iii) If firm i has no cost advantage at x, that is, $m_i(x) \geq m_j(x)$, then for any price $p_i \geq m_i(x)$ the firm has no demand when $m_i(x) > m_j(x)$ since firm j can always undercut p_i, or it makes no profit when $m_i(x) = m_j(x)$. In any case, firm i earns zero profits and pricing at $m_i(x)$ is optimal.

Thus, as is often suggested in the marketing literature, $p_j^*(x)$ is set on the basis of either demand considerations, or, a combination of demand and competition considerations, or finally, purely cost considerations; the dominant consideration being determined by the consumer location x.

The above-described equilibrium is the only one of the game. Intuitively, this is easy to understand. Indeed, for a given location x in which no firm has a monopoly position, Bertrand competition drives prices down to the level of the larger total marginal cost, $\max\{m_1(x), m_2(x)\}$, which then allows the firm with the cost advantage, say firm i, to charge $\bar{p}_i(x)$. If firm i has a monopoly position then $p_i^M(x)$ necessarily obtains; and similarly for firm j.[9]

The *equilibrium market price schedule* is then

$$p^*(x) = \min\{p_1^*(x), p_2^*(x)\} \text{ for all } x \in X.$$

That is, $p^*(x)$ equals the minimum of the monopoly prices, $p_i^M(x)$, the constrained prices, $\bar{p}_i(x)$, and the maximum of the total marginal costs, $\max\{m_1(x), m_2(x)\}$.

The equilibrium is illustrated in the following example.

The set X is a linear segment of length 1; firms 1 and 2 are located respectively at $y_1 = 0$ and $y_2 = 1$; consumers have linear demand functions $1 - p$; production costs are zero and transportation costs are given by tx. For $1/2 \leq t \leq 2$, it is then easy to verify that

$$p_1^*(x) = p_1^M(x) = \frac{1}{2} + \frac{t}{2}x \qquad \text{for } 0 \leq x \leq (2t-1)/3t,$$

$$p_1^*(x) = m_2(x) = t(1-x) \qquad \text{for } (2t-1)/3t \leq x \leq \frac{1}{2},$$

and $$p_1^*(x) = m_1(x) = tx \qquad \text{for } \frac{1}{2} < x \leq 1.$$

A similar argument can be developed for $p_2^*(\cdot)$. Figure 1 represents the resulting market price schedule.

Interestingly, we notice that, over a certain range of locations ($(2t-1)/3t \leq x \leq \frac{1}{2}$ in Figure 1), the equilibrium price schedule is *decreasing* in the distance to the firm. The reason is that a firm faces fiercer competition in remote places—and thus sets lower delivered prices—than it does at home, for it has to compete with another firm which is closer to these points. This seems to be confirmed by the observations made by Greenhut (1981) for whom negatively sloped delivered price schedules frequently characterize oligopolistic firms in Japan and West Germany.[10]

---

[9] A more technical argument could be given following the lines of the proof of Theorem 1 in P. J. Lederer and A. P. Hurter (1986), (who consider the case of inelastic local demands).

[10] Greenhut reports that "this type of pricing...was never found...in the United States. Instead, the American firms always charged higher prices to more

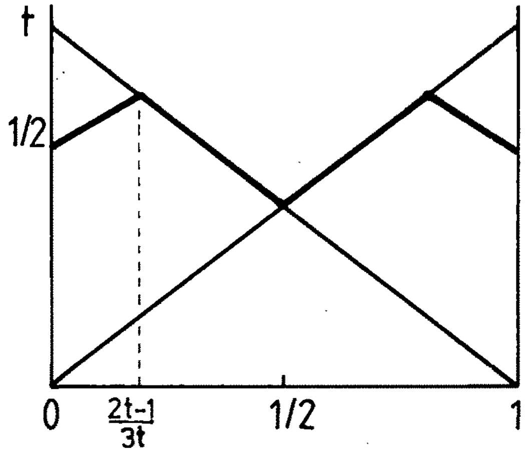


FIGURE 1. THE HEAVY LINE IS THE EQUILIBRIUM MARKET PRICE SCHEDULE

Proposition 1 states our result.

PROPOSITION 1: *Assume that the intersection of the potential market areas of the firms is a nonnegligible set. If there is competition in price schedules, then the unique Nash equilibrium market price schedule is*

$$p^*(x) = \min\{p_1^*(x), p_2^*(x)\}.$$

Obviously, equilibrium prices involve price discrimination. In other words, *uniform pricing is never an equilibrium* when the potential market areas have a nonnegligible intersection. If this intersection were to be negligible, then the two firms would become spatial monopolists. For one particular class of demand functions, negative exponential, discriminatory pricing boils down to uniform pricing.

III. Commitment to a Price Policy

In this section we consider two-stage games in which firms may commit to a particular price policy, uniform pricing (U), or may not commit at all and stay free to choose an unrestricted price schedule (D) at the market competition stage (this involves price discrimination in general).[11]

In order to make the analysis tractable, we will restrict ourselves to simple cases in which consumers are uniformly distributed over some space X (a segment or a circle) and each one of them has an *inelastic* demand for one unit of the good. Attention will be focused on subgame-perfect equilibria of the two-stage games in which firms anticipate the resulting Nash equilibria in prices at the second stage when choosing its price policy at the first stage. If one firm chooses uniform pricing and the other discriminatory pricing, there may not be a simultaneous move Nash equilibrium (in pure strategies) at the second stage. In what follows, we will assume that the firm which chooses to price uniformly will move first and be the price leader while the other firm will react optimally to the leader's price. This situation fits the single-basing point pricing (BPP) system in spatial price competition and we will see that it has a natural interpretation in the context of product differentiation. We deal first with spatial price competition and the equilibrium analysis of the BPP system, and second with product differentiation and the pricing of varieties.

A. *Spatial Price Competition and Basing Point Pricing*

We suppose now that consumers are uniformly distributed with a unit density on the interval $X = [0,1]$, with firm 1 located at 0 and firm 2 at 1. Firms have constant marginal production costs and, without loss of generality, let them be 0 and $c \geq 0$ for firms 1 and 2, respectively. Transportation costs are linear with slope t ($t > 0$).

---

distant buyers, or at the limit followed uniform delivered prices over their market space. Quite conceivably it is the Robinson-Patman Act that causes the delivered price patterns of American firms to differ from those of firms in West Germany and Japan" (see Greenhut, 1981, p. 84). This suggests that the corresponding institutional constraint is binding for many American firms. For an illustration, see the simulation made by B. F. Hobbs (1986a), of a deregulated bulk power market in the United States.

[11] A model similar in spirit is developed by N. Singh and X. Vives (1984).

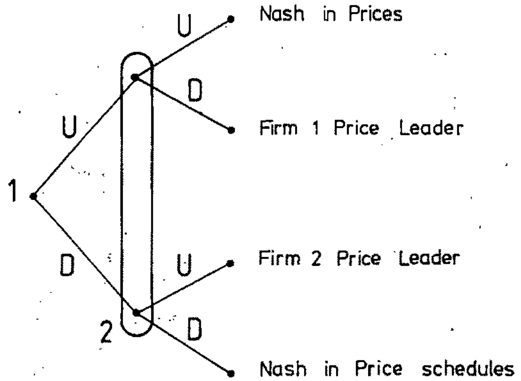


FIGURE 2. GAME TREE FOR THE SEQUENTIAL GAME WHERE FIRMS MAY EITHER COMMIT TO PRICE UNIFORMLY (U) OR KEEP THEIR FREEDOM OF PRICING (D)

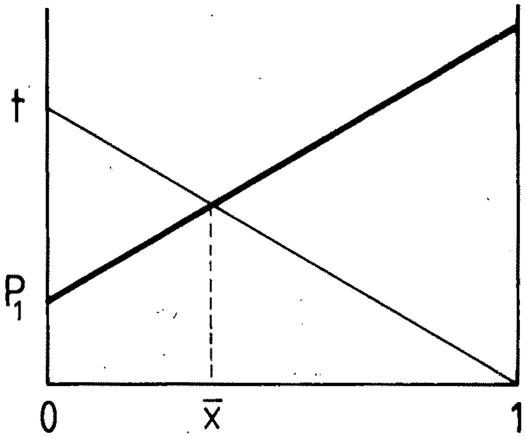


FIGURE 3. SINGLE-BASING POINT PRICING WITH BASE AT 0. THE HEAVY LINE IS THE FULL MARKET PRICE $p_1 + tx$. THE INTERSECTION WITH THE TOTAL MARGINAL COST OF FIRM 2 DETERMINES $\bar{x}$, THE MARKET SHARE OF THE LEADER

Firms may commit to uniform pricing first and then compete in prices accordingly. Four possible cases may arise: (U,U), (U,D), (D,U), or (D,D). If both firms choose to price uniformly, (U,U), then a Nash equilibrium in (FOB) prices obtains at the second stage. If both firms choose to keep complete freedom of pricing, (D,D), then a Nash equilibrium in price schedules obtains at the second stage. If firm 1 chooses to price uniformly and firm 2 does not commit, (U,D), then firm 1 will be a price leader and firm 2 will react optimally to its price (Figure 2 depicts the game tree).

This is a natural *competitive* view of the BPP system with a single-basing point (in which firm 1 is located at $y_1 = 0$). The leader announces a uniform base price, and the price the consumer pays is just the base price plus the transportation cost from 0 to the location of the consumer no matter what firm serves the consumer. Given the base price set by the leader, firm 2 then just undercuts the corresponding full price wherever possible. In other words, the market area of firm 2 is defined by the set of locations for which firm 1's full price is larger than the marginal production and transportation cost of firm 2. This is so because firm 1 is the first mover so that, for any posted price p_1, firm 2 can always capture the demand on $\{x \in X;\ P_1 + t|y_1 - x| > c + t|y_2 - x|\}$ by selling to consumers at $x\,\varepsilon$-below firm 1's full price (see Figure 3 where the market areas of firms 1 and 2 are given by $[0,\bar{x}]$ and $[\bar{x},1]$, respectively, when $c = 0$).[12] Firm 1 is, therefore, a price leader that sets a uniform FOB price and firm 2 reacts optimally to the leader's base price by price discriminating. The case (D,U) is similar. We have thus that the mixed cases (U,D) and (D,U) represent single BPP equilibria with base points 0 and 1, respectively. Our approach results in separated market areas and, therefore, no cross-hauling occurs, that is, there is no location where two-way trade is observed.[13]

We would like to examine the stability of basing point pricing in our competitive context. In particular, we would like to know whether BPP can emerge as an equilibrium of our two-stage game.

---

[12] Notice that the way demand is allocated between firms in the event of a price tie has been slightly modified with respect to the convention made in Section III where it was assumed that customers were assigned to the firm with the cost advantage.

[13] In practice, however, market areas are not so clearly delineated as in our competitive interpretation of BPP and cross-hauling may occur. This observation already suggests that the actual implementation of BPP may not be consistent with vigorous price competition since cross-hauling implies that firm 2 does not fully exploit its cost advantage with respect to the price set by firm 1.

In order to find the subgame-perfect equilibria of our game we compute the firms' payoffs in each of the four possible situations. The cases (U,U) and (D,D) are standard and we will review them very briefly. We will assume that $t > c$ to ensure that firm 2 (the high-cost firm) is not priced out of the market.

(U,U): The market boundary for the two firms is given by the location $\bar{x}$ of the consumer who is indifferent between buying from either firm: $p_1 + t\bar{x} = p_2 + t(1-\bar{x})$, from which it immediately follows that $\bar{x} = (p_2 - p_1 + t)/2t$. As consumers are distributed with a unit density, profits of firm 1 are given by $\pi_1 = p_1\bar{x}$ and those of firm 2 by $\pi_2 = (p_2 - c)(1-\bar{x})$. The unique pair of equilibrium prices is obtained from the first-order conditions as

$$\left(t + \frac{c}{3}, t + \frac{2c}{3}\right),$$

yielding market areas

$$\left(\frac{1}{2} + \frac{c}{6t}, \frac{1}{2} - \frac{c}{6t}\right)$$

and equilibrium profits

$$\left(\frac{1}{2t}\left(t + \frac{c}{3}\right)^2, \frac{1}{2t}\left(t - \frac{c}{3}\right)^2\right).$$

(D,D): If both firms compete in price schedules, the equilibrium price schedule $p^*(x)$ is $\max\{tx, c + t(1-x)\}$, with $x \in [0,1]$, since localized Bertrand competition drives prices down to the higher total marginal cost (see Proposition 1). The market boundary is given by $x^* = 1/2 + c/2t$, while equilibrium profits are

$$\pi_1 = \int_0^{x^*} [c + t(1-x) - tx]\, dx$$

$$= \frac{1}{4t}(c+t)^2 \quad \text{for firm 1, and}$$

$$\pi_2 = \int_{x^*}^1 [tx - (c + t(1-x))]\, dx$$

$$= \frac{1}{4t}(t-c)^2 \quad \text{for firm 2.}$$

(U,D): Here the efficient firm is the leader and prices uniformly at p_1. The market boundary $\bar{x}$ is determined by $p_1 + t\bar{x} = c + t(1-\bar{x})$, which yields $\bar{x} = (t + c - p_1)/2t$, since the optimal response of firm 2 is to match firm 1's full price $p_1 + tx$, whenever possible, that is when $p_1 + tx \geq c + t(1-x)$. Profits of firm 1 are given by $\pi_1 = p_1\bar{x}$ and the optimal price for firm 1 is $p_1{}^* = t + c/2$ with associated market boundary $x^* = t + c/4t$, yielding profits of $(t+c)^2/8t$. The equilibrium price schedule of firm 2 is $p_2^*(x) = \max\{p_1^* + tx, c + t(1-x)\}$ and the equilibrium profits are

$$\pi_2 = \int_{x^*}^1 \left[\frac{t+c}{2} + tx - (c + t(1-x))\right] dx$$

$$= \frac{(3t-c)^2}{16t}.$$

(D,U): We have now that the inefficient producer (firm 2) prices uniformly and is the price leader. A symmetric argument yields a market boundary $\bar{x} = p_2 + t/2t$, an optimal price for firm 2 equal to $t + c/2$ with associated equilibrium profits $(t-c)^2/8t$. Equilibrium profits for firm 1 are $(3t+c)^2/16t$. Table 1 summarizes the payoffs for the firms.

It is clear from the table that to keep pricing freedom and to price discriminate is a dominant strategy for any firm no matter the difference in the production costs of the firms (provided that $c \leq t$). Commiting to uniform pricing is, therefore, a dominated choice. Proposition 2 states the result.

PROPOSITION 2: *In the sequential commitment game, choosing the price discrimination policy is a dominant strategy for any firm and, consequently,* (D, D) *with resulting market price schedule* $p^*(x) = \max\{tx, c + t(1-x)\}$ *is the unique subgame-perfect equilibrium.*

Proposition 2 says that no firm, not even the more efficient one, wants to be the price leader taking as basing point its location and, therefore, single BPP is not a stable configuration since it is not an equilibrium of our two-stage game. This suggests the

TABLE 1—SUMMARY OF FIRMS' PAYOFFS

1 \ 2	U	D
U	$\frac{1}{2t}\frac{(3t+c)^2}{9},\ \frac{1}{2t}\frac{(3t-c)^2}{9}$	$\frac{1}{2t}\frac{(t+c)^2}{4},\ \frac{1}{2t}\frac{(3t-c)^2}{8}$
D	$\frac{1}{2t}\frac{(3t+c)^2}{8},\ \frac{1}{2t}\frac{(t-c)^2}{4}$	$\frac{1}{2t}\frac{(t+c)^2}{2},\ \frac{1}{2t}\frac{(t-c)^2}{2}$

hypothesis that BPP cannot be explained in the context of a noncooperative model in which firms can choose their price policy and in which there is no repeated competition. Our analysis thus contrasts with the view (see, for example, D. H. Haddock, 1982) that BPP is fundamentally "competitive" and suggests that theoretical explanations of BPP should consider its role as a coordinating and collusive device (see Scherer, 1980, and G. Stigler, 1949). Nevertheless, our hypothesis will obviously not hold when BPP coincides with the discriminatory solution. This would be the case, for example, when firm 1 has a second plant established next to firm 2 at location $y_2 = 1$ and both firms are equally efficient. Indeed, in the region where no firm has a cost advantage, competition would drive prices down to total marginal cost, while in the region in which firm 1 has a cost advantage, the delivered price will be equal to the total marginal cost of firm 2, that is, the equilibrium market price schedule would be given by $t(1-x)$ for $0 \leq x \leq 1$. It is then clear that the above solution corresponds to BPP with a single-basing point at $y_2 = 1$.

In the symmetric case where firms are of equal productive efficiency, $c = 0$, we have a typical Prisoner's Dilemma situation since to price discriminate is a dominant strategy but firms would make more profits by pricing uniformly.[14] In the (U,U)-case each firm earns $t/2$, whereas in the (D,D)-case each firm earns $t/4$. Consumer surplus is larger in this latter case since the full price at x is given by $p^D(x) = \max\{tx, t(1-x)\}$, whereas in the (U,U)-case the uniform price is t and, therefore, the full price at x is given by $p^U(x) = \min\{t + tx, t + t(1-x)\}$, which is strictly larger than $p^D(x)$ for $0 < x < 1$: each consumer in $]0,1[$ is, therefore, better off under discriminatory than uniform pricing while consumers located at $x = 0$ and $x = 1$ are indifferent (see Figure 4). Note nevertheless that total surplus is equal in both cases; the pricing policy only affects the distribution of surplus between firms and consumers. To check our claim suppose that consumers have a (high enough) reservation price v ($v \geq 3t/2$) and that the utility provided by the consumption of one unit of the good equals v minus the full price paid by the consumer. We have then that total surplus in the market is $v - t/2$ in both cases, the differences between the reservation price and the transportation cost of the consumer located in the middle of the segment.

Imagining that location 0 corresponds to Pittsburgh and location 1 to Chicago, the case (U,D) would represent the Pittsburgh-plus system implemented until the mid-1920s in the U.S. steel industry. After that period, several basing points were introduced involving, among others, both Pittsburgh and Chicago. Superficially, the (U,U)-case could then be viewed as a very simplified version of the multiple-basing point system with two basing points at locations 0 and 1. Actually, this turns out not to be true as this system is always associated with the so-called "alignment rule" (see Phlips, 1983, ch. 1, and Scherer, 1980, ch. 11). According to this rule, a firm will always accept to sell at the lower

---

[14] This result holds in various models: linear demand and endogenous location (A. Kats, 1987) and product differentiation modeled by the logit (S. P. Anderson et al., 1987). Notice, however, that discriminatory pricing may yield higher profits than uniform pricing when firms have monopoly positions in some large enough segments of the market.

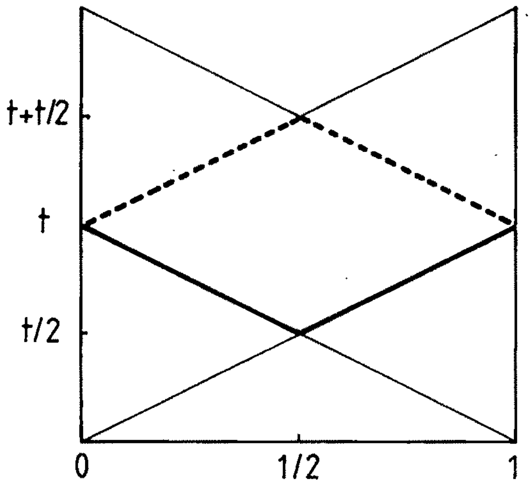


Figure 4. The Heavy Line is the Market Equilibrium Price Schedule with Discriminatory Pricing, $p^D(x) = \max\{tx, t(1-x)\}$. The Dotted Line Corresponds to the Uniform Pricing Case, $p^U(x) = \min\{t(1+x), t(2-x)\}$.

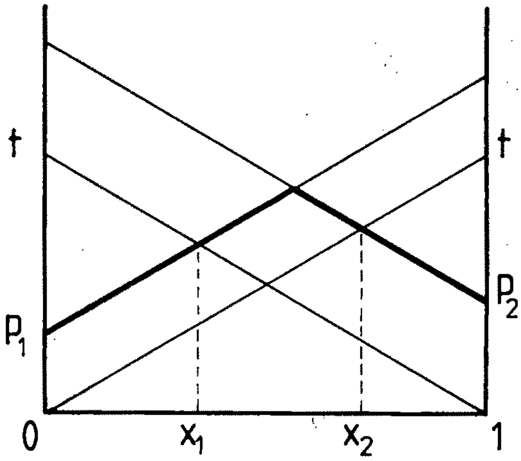


Figure 5. Multiple-Basing Point Pricing with Bases at 0 and 1. The Heavy Line Represents the Delivered Prices Faced by Consumers

full price so that firms quote the same delivered price to everyone at each location (provided that this price is large enough for the firm to cover its marginal cost). Let us assume, for simplicity, that $c = 0$ so that both firms are equally efficient. Then, if firms 1 and 2 charge base prices p_1 and p_2, we can see in Figure 5 that customers located in $[x_1, x_2]$ may pass orders to either firm since, by assumption, they both set the same delivered price at $x \in [x_1, x_2]$. As any assignment rule is a priori arbitrary at this stage of the analysis, we may follow A. Smithies (1942) and suppose that firms equally share the local market in the interval $[x_1, x_2]$. In this case, it is readily verified that, at the Nash equilibrium in base prices, both firms charge a base price that is just equal to the common equilibrium price arising in the (U, U)-case, that is, $p_1^* = p_2^* = t$. However, because of the existence of cross-hauling over $[x_1, x_2]$, the corresponding equilibrium profits are lower and given by $\frac{3}{8}t$. It is interesting to observe that these profits are still larger than those earned at the noncooperative equilibrium of our two-stage game (D, D), *but* lower than those obtained at the "cooperative" solution (U, U). What Smithies' approach leaves unexplained is the market sharing in the interval $[x_1, x_2]$ since, according to our results, firms have a strong incentive to price discriminate. Market sharing in the common area $[x_1, x_2]$ gives the multiple BPP a noncompetitive flavor.[15]

B. *Product Differentiation and the Pricing of Varieties*

Consider consumers distributed uniformly over the unit circle with firms 1 and 2 in arbitrary locations, the shortest arc distance between them being s (that is, $s \leq 1/2$). In Figure 6, firm 1 is located at $x = 0$ and firm 2 at $x = s$. Suppose that firms have no production costs and that transportation costs are quadratic with coefficient t and no linear term. (As explained in Section I, here transportation costs will be the costs associated with transforming a base product into a variety or specialized product.)

---

[15] However, Smithies is probably right when he claims that this system is far from complete agreement since profits are lower than in the second-best (U,U)-case.

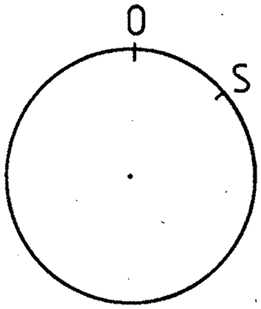


FIGURE 6. FIRM 1 IS LOCATED AT $x = 0$, FIRM 2 AT $x = s$

In the situation we have in mind each firm produces a base product, corresponding to its location in product space (the circle), and at the first stage it has to decide whether to (potentially) offer the whole array of varieties and price discriminate or just to offer the base product and price uniformly therefore. In the latter case the supply of specialized varieties is left to outside independent producers which, we will assume, price competitively at cost. In any case, we assume that the costs of redesigning the base product increase quadratically with the distance at which the variety chosen is located. The first-stage choice could also be interpreted as a decision as whether to integrate forward or not. The base product could be a base chemical or steel and the varieties the specialized chemicals or steels. Whatever the interpretation, if a firm chooses to produce only the base product, then it will choose a uniform price, and if it chooses to offer the array of varieties, it will choose a price schedule when the market stage comes. As before we assume that in the mixed cases (U, D) or (D, U) the firm that prices uniformly moves first and therefore Figure 2 represents the game tree of our game. This may come about because this firm may need to advertise its price in order to get any sales (put advertisements in the newspapers, for example), whereas the price-discriminating firm may just announce that it will meet the competition (in any case to announce the whole price schedule in an intelligible way may be too complex and costly due to the large number of varieties).

TABLE 2—FIRM 1'S PROFITS

1 \ 2	U	D
U	$\frac{1}{2}s(1-s)t$	$\frac{1}{8}s(1-s)t$
D	$\frac{9}{16}s(1-s)t$	$\frac{1}{4}s(1-s)t$

It is a simple exercise to find the second-stage equilibria for the different cases. Note that since transportation costs are quadratic (with no linear term) an equilibrium in prices will always exist in the (U, U)-case whatever the locations of the firms. In the Appendix we provide the computations that lead to the profits of firm 1 in the four cases given in Table 2 (profits of firm 2 have similar expressions).

As before, to choose the price-discriminating strategy is dominant for any firm in the circle model with quadratic costs. No firm wants to produce only the base product whatever the rival firm strategy is. It always pays to offer the whole array of varieties and let the market decide which ones will be effectively produced by each firm. As in the previous models, price discrimination gives more flexibility to a firm to respond to any potential strategy of the rival. A product proliferation strategy (in terms of horizontal differentiation) can thus be seen as an attempt by a firm to secure a flexible position for the ensuing price competition.

In fact we have checked that the same is true in a unit-segment model with quadratic transportation costs if the firms are symmetrically located. The payoffs to firm 1 are given by Table 2's dropping the common terms $s(1-s)t$. If firms are asymmetrically located in the segment, then (D, D) is still the unique subgame-perfect equilibrium of the two-stage game but it is no longer necessarily a dominant strategy equilibrium. Going back to Table 2, we observe that, as in the location model considered in Section III, Part A, the firms face a Prisoner's Dilemma-type situation with higher profits in the (U, U)-case whereas the equilibrium dictates (D, D). Total surplus is equal in both situations and therefore, consumer surplus is higher with price discrimination and the two firms (potentially) offering the whole array of varieties.

IV. Concluding Remarks

We have examined the implications of letting firms choose their price policy in the context of a spatial competition model with given locations for firms. Either firms choose simultaneously price policy and actual prices, or firms may commit to a certain policy (uniform pricing) before the actual price competition takes place. The general conclusion is that *there is a robust tendency for a firm to choose the discriminatory policy* since it is more flexible and does better against any generic strategy of the rival, although, as we have seen in the models of Section III, firms may end up worse off than if they choose to price uniformly.

Furthermore, in those models, prices that consumers paid under uniform pricing were higher than under discriminatory pricing. This is not totally surprising: denying a firm the right to meet the price of a competitor on a discriminatory basis provides the latter with some protection against price attacks. The effect is then to weaken competition, contrary to the belief of the proponents of naive application of legislation prohibiting price discrimination like the Robinson-Patman Act in the United States, or similar recommendations of the Price Commission in the United Kingdom. Actually, as observed by E. M. Hoover,

> The difference between market competition under FOB pricing (with strictly delineated market areas) and under discriminatory delivered pricing is something like the difference between trench warfare and guerilla warfare. In the former case all the fighting takes place along a definite battle line; in the second case the opposing forces are intermingled over a broad area.
> [1948, p. 57]

Our results are short-run results since both the location and the number of firms are given. We know that the choice of a particular price policy leads to different long-run equilibrium patterns (see, for example, Greenhut et al., 1987, Part III). When policy and prices are chosen simultaneously, Proposition 1 (that can be generalized to the case of any number of firms) indicates that, at the long-run equilibrium, firms will choose to price discriminate in the absence of institutional constraints. Thus, at the long-run equilibrium, uniform pricing would not be observed. P. J. Lederer and A. P. Hurter (1986) have shown that, with perfectly inelastic demand, two price-discriminating firms will locate in order to minimize total transportation costs. In a free-entry context, the configuration of firms minimizing total production and transportation costs corresponds to a long-run equilibrium (see W. B. MacLeod et al., 1985). Furthermore, in a free-entry, zero-profit equilibrium with linear demand, B. F. Hobbs (1986b) shows that welfare under discriminatory pricing is higher than under mill pricing for a wide range of fixed cost values.[16] These results, together with ours, point to the social desirability of price discrimination in spatial competitive markets. However, more work is called for before having robust policy-oriented recommendations.

Our analysis has also helped to clarify some issues related to standard business practices. With respect to geographical pricing policies, the tendency toward price discrimination makes in general the basing point system unstable since no firm, not even a very efficient one, would like to commit to uniform (FOB) pricing. Therefore, if the system is observed in practice, most probably it is because it serves as a coordinating or collusive device in a situation of repeated competition. With respect to pricing in a product-differentiation context, the link between pricing policies and product strategies has been highlighted, noticing that for a firm to be able to price discriminate it must offer a band of varieties. Since it always pays to price discriminate, this may explain in part the observed change of firms' strategies from producing a single-standarized product to offering a whole spectrum of options.

---

[16] Like most models of spatial monopolistic competition, Hobbs assumes that firms are equally spaced. However, no justification is provided for that property.

Much work needs to be done to gain a solid understanding of current business pricing policies. Ours is only a first attempt and an illustration of how simple game-theoretic techniques may help to illuminate the issues involved.[17]

APPENDIX

Assume that locations (x) along the circle of unit length are measured in a trigonometric manner. Firm 1 is at $x = 0$ and firm 2 at $x = s$ with $0 < s \leq 1/2$.

(U, U): Consumers are indifferent between purchasing from either firm at points $\bar{x} = (p_2 - p_1 + ts^2)/2ts$ and $\bar{y} = p_1 - p_2 + t(1 - s^2)/2t(1-s)$. Consequently, profits of firm 1 are $\pi_1 = p_1(1 - \bar{y} + \bar{x})$ and profits of firm 2 are $\pi_2 = p_2(\bar{y} - \bar{x})$.The unique price equilibrium is given by $(ts(1-s), ts(1-s))$, yielding equilibrium profits $(ts(1-s)/2, ts(1-s)/2)$.

(D, D): If both firms price discriminate, it is readily verified that the equilibrium price schedule is given by

$$
\text{(A1)}\quad p^*(x) = \begin{cases} \max\{tx^2, t(x-s)^2\} \\ \quad \text{for } 0 \leq x \leq 1/2 \\ \max\{t(1-x)^2, t(x-s)^2\} \\ \quad \text{for } 1/2 \leq x \leq 1/2 + s \\ \max\{t(1-x)^2, t(1-x+s)^2\} \\ \quad \text{for } 1/2 + s \leq x \leq 1. \end{cases}
$$

The market boundaries are given by $\bar{x} = s/2$ and $\bar{y} = 1/2 + s + s/2$, whereas the equilibrium profits are

$$
\text{(A2)}\quad \pi_1 = \int_0^{s/2} [p^*(x) - tx^2]\,dx + \int_{1/2+s}^{1} [p^*(x) - t(1-x)^2]\,dx = \frac{ts(1-s)}{4};
$$

similarly,

$$
\text{(A3)}\quad \pi_2 = \frac{ts(1-s)}{4}.
$$

(U, D): Firm 1 is the leader and prices uniformly at p_1. Given p_1, firm 2 sets a price at x which is equal to the maximum of firm 1's full price at x and firm 2's transportation cost to x. Accordingly, the market boundaries are as follows: (i) for $p_1 \leq ts^2$, we have $\bar{x} = (-p_1 + ts^2)/2ts$ and $\bar{y} = p_1 + t(1 - s)^2/2t(1-s)$; ($ii$) for $ts^2 < p_1 \leq ts(1-s)$, we have $\bar{\bar{x}} = -p_1 + ts^2 + 2ts)/2ts$ and $\bar{\bar{y}} = \bar{y}$ (when $p_1 > ts(1-s)$ and all consumers buy from firm 2). Firm 1's profits are respectively defined by $\pi_1 = p_1(1 - \bar{y} + \bar{x})$ and $\pi_1 = p_1(\bar{\bar{x}} - \bar{y})$. $A x$ $\bar{\bar{x}} = \bar{x} + 1$, the expression of the profit function is uniquely determined and the profit-maximizing price of firm 1 is $p_1{}^* = ts(1-s)/2$. The corresponding demand is $1/4$ and the equilibrium profits of firm 1 are $ts(1-s)/8$. For firm 2, the equilibrium price schedule is given by

$$
\text{(A4)}\quad p_2{}^*(x) = \begin{cases} \max\{p_1{}^* + tx^2, t(x-s)^2\}, \\ \quad \text{for } 0 \leq x \leq 1/2 \\ \max\{p_1{}^* + t(1-x)^2, t(x-s)^2\}, \\ \quad \text{for } 1/2 \leq x \leq 1/2 + s \\ \max\{p_1{}^* + t(1-x)^2, t(1-x+s)^2\}, \\ \quad \text{for } 1/2 + s \leq x \leq 1 \end{cases}
$$

and the equilibrium profits are

$$
\text{(A5)}\quad \pi_2 = \int_0^{\bar{x}} [p_2{}^*(x) - tx^2]\,dx + \int_{\bar{y}}^{1} [p_2{}^*(x) - t(1-x)^2]\,dx
$$

---

[17] Many other relevant issues have been left out. For example, we have considered complete information models (for models with informed and uninformed buyers, see, for example, S. Salop and J. Stiglitz, 1977, and M. L. Katz, 1984) and we have ignored the possibility of nonlinear pricing (see D. Spulber, 1981, for a study of nonlinear pricing in a spatial context).

or

$$\pi_2 = \int_{\bar{\bar{y}}}^{\bar{\bar{x}}} \left[p_2^*(x) - t(1-x)^2 \right] dx,$$

which are both equal to

$$\tfrac{9}{16} ts(1-s).$$

(D, U): This is perfectly similar to the above one (up to a rotation of $1-s$ and a permutation of firms' names). Equilibrium profits are therefore

$$\pi_1 = \frac{9}{16} ts(1-s) \quad \text{and} \quad \pi_2 = \frac{ts(1-s)}{8}.$$

REFERENCES

Anderson, S. P., de Palma, A. and Thisse, J.-F., "Spatial Price Policies and Duopoly," Université Catholique de Louvain, CORE Discussion Paper No. 8737, 1987.

Augier, L. and Icikovics, J.-P., "L'automobile 1982–1990," *Science et Vie*, June 1982, Special Issue.

Beckmann, M. J., "Spatial Price Policies Revisited," *Bell Journal of Economics*, Autumn 1976, *7*, 619–30.

Greenhut, J. and Greenhut, M. L., "Spatial Price Discrimination, Competition and Locational Effects," *Economica*, November 1975, *42*, 401–19.

______, ______, and Li, S. Y., "Spatial Pricing Patterns in the United States," *Quarterly Journal of Economics*, March 1980, *94*, 329–50.

Greenhut, M. L., "Spatial Pricing in the United States, West Germany and Japan," *Economica*, February 1981, *48*, 79–86.

______, Norman, G. and Hung, C.-S., *The Economics of Imperfect Competition. A Spatial Approach*, Cambridge: Cambridge University Press, 1987.

______ and Ohta, H., "Monopoly Output Under Alternative Spatial Pricing Techniques," *American Economic Review*, September 1972, *62*, 705–13.

Haddock, D. H., "Basing-Point Pricing; Competitive vs. Collusive Theories," *American Economic Review*, June 1982, *72*, 289–306.

Hamilton, J., Thisse, J.-F. and Weskamp, A., "Spatial Discrimination: Bertrand vs. Cournot in a Model of Location Choice," mimeo., Department of Economics, University of Florida, 1987.

Hobbs, B. F., (1986a) "Network Models of Spatial Oligopoly with an Application to Deregulation of Electricity Generation," *Operations Research*, May-June 1986, *34*, 395–409.

______, (1986b) "Mill Pricing vs. Spatial Price Discrimination Under Bertrand and Cournot Spatial Competition," *Journal of Industrial Economics*, December 1986, *35*, 173–91.

Holahan, W. L., "The Welfare Effects of Spatial Price Discrimination," *American Economic Review*, June 1975, *65*, 498–503.

Hoover, E. M., *The Location of Economic Activity*, New York: McGraw-Hill, 1948.

Kats, A., "Price Discrimination in a Spatial Model of Monopolistic Competition with Downward Sloping Demand," mimeo., Department of Economics, Virginia Polytechnic Institute and State University, 1987.

Katz, M. L., "Price Discrimination and Monopolistic Competition," *Econometrica*, November 1984, *52*, 1453–71.

Lederer, P. J. and Hurter, A. P., "Competition of Firms: Discriminatory Pricing and Location," *Econometrica*, May 1986, *54*, 623–40.

Love, R. F. and Morris, J. G., "Mathematical Models of Road Travel Distances," *Management Science*, February 1979, *25*, 130–39.

MacLeod, W. B., Norman, G. and Thisse, J.-F., "Price Discrimination and Equilibrium in Monopolistic Competition," Université Catholique de Louvain, CORE Discussion Paper No. 8506, 1985.

Norman, G., "Spatial Pricing with Differentiated Products," *Quarterly Journal of Economics*, May 1983, *97*, 291–310.

Oi, W. Y., "A Disneyland Dilemma: Two-Part Tariffs for a Mickey-Mouse Monopoly," *Quarterly Journal of Economics*, February 1971, *85*, 77–96.

Phlips, L., *The Economics of Price Discrimination*, Cambridge: Cambridge University Press, 1983.

Pigou, A. C., *The Economics of Welfare*, London: Macmillan, 1920.

Robinson, J., *The Economics of Imperfect Competition*, London: Macmillan, 1933.

Salop, S. and Stiglitz, J., "Bargain and Ripoffs: A Model of Monopolistic Competition with Price Dispersion," *Review of Economic Studies*, October 1977, *44*, 493–510.

Scherer, F. M., *Industrial Market Structure and Economic Performance*, Chicago: Rand McNally, 1980.

Schmalensee, R., "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, March 1981, *71*, 242–47.

Schwartz, M., "The Perverse Effects of the Robinson-Patman Act," *Antitrust Bulletin*, Fall 1986, *31*, 733–57.

Singh, N. and Vives, X., "Price and Quantity Competition in a Differentiated Duopoly," *The Rand Journal of Economics*, Winter 1984, *15*, 546–54.

Sloan, A. P., Jr., "*My Years with General Motors*," in J. McDonald with C. Stevens, ed., New York: Doubleday, 1972.

Smithies, A., "Monopolistic Price Policy in a Spatial Market," *Econometrica*, January 1941, *9*, 63–73.

______, "Aspects of the Basing-Point System," *American Economic Review*, December 1942, *32*, 705–26.

Spulber, D. F., "Spatial Nonlinear Pricing," *American Economic Review*, December 1981, *71*, 923–33.

Stigler, G. J., "A Theory of Delivered Price Systems," *American Economic Review*, December 1949, *39*, 1143–59.

Ward, J. E. and Wendell, R. E., "Measuring Distances via Block Norms with an Application to Facility Location Models," *Operations Research*, September-October 1985, *33*, 1074–90.

Price Commission, (1978a) "UG Glass Containers—Price of Glass Containers," Price Commission Report No. 5, House of Commons 170, H.M.S.O, 1978.

Price Commission, (1978b) "Metal Box Ltd.—Open Top Food and Beverage and Aerosol Cans," Price Commission Report No. 3, House of Commons 135, H.M.S.O., 1978.

Price Commission, (1978c) "Tate and Lyle Refineries Limited—Sugar and Syrup Products," Price Commission Report No. 6, House of Commons 224, H.M.S.O, 1978.

Will Share Contracts Increase Economic Welfare?

By RUSSELL COOPER*

This paper develops a two-sector model with imperfect competition in order to explore the positive and normative aspects of the share system advanced by Weitzman in his book. In this model, the degree of competitiveness is measured by a parameter $\eta = [(F-1)/F]$, *where* F *is the number of firms in each sector. With fixed-wage contracts, sector-specific shocks generate aggregate fluctuations in employment and output through a multiplier effect. Introducing share contracts will not yield a Pareto-dominant allocation unless the share parameter is exactly equal to* η. *When the share parameter equals* η, *welfare is increased. Thus to reap the benefits of introducing share contracts requires rather exact information on the competitiveness of the economy.*

In a series of recent publications, Martin Weitzman (1983–1985) has strongly advocated the construction of a system of share contracts to mediate transactions in the labor market. Under such a system, workers' compensation would automatically fall with increases in employment. These contracts, Weitzman argues, would provide the correct incentives at the microlevel, so that workers and firms would act to reduce macroeconomic fluctuations. Weitzman (1984) says that,

> The lasting solution to stagflation requires going inside the workings of a modern capitalist economy and correcting the underlying structural flaw directly at the level of the individual firm by changing the nature of labor remuneration. An alternative payment system where it is considered perfectly normal for a worker's pay to be tied to an index of the firm's performance, say a share of its revenue or profits, puts in place exactly the right incentives automatically to resist unemployment and inflation. [Weitzman, p. 3]

The response to this challenging proposal has been mixed. At one extreme, The *New York Times*, in an editorial entitled, "Best Idea Since Keynes," extolled a number of virtues associated with Weitzman's proposal. The *Times* warned that "ideas that promise so much usually succumb to general skepticism," and noted that "the idea needs testing for analytic errors and practical examination..."

The present paper is an attempt in that direction and analyzes an explicit model economy for the purposes of comparing and contrasting the positive and normative properties of the two compensation schemes highlighted in Weitzman's work: a fixed-wage system and a share system.[1] The basic struc-

---

*Department of Economics, College of Business Administration, University of Iowa, Iowa City, Iowa 52242. This paper draws on a series of conversations with colleagues at Yale University as well as the spirited discussion during a conference held at Yale on Martin Weitzman's *The Share Economy*. The conference proceedings are contained in Andrew John and William Nordhaus (1986). I am grateful to Andrew John, Bill Nordhaus, Bruce Smith, Martin Weitzman, John Taylor, and anonymous referees for helpful comments and suggestions on this paper. I thank seminar participants at the 1985 NBER Summer Institute, the NBER Conference on "Keynesian and Classical Economics After 50 Years" and Purdue University for comments. All errors in analysis and interpretation are mine alone.

[1] Weitzman (1985) also seeks to provide a more detailed discussion of the merits of a share system. Our papers share the goal of attempting to further explicate Weitzman's (1984) arguments for a share economy. The models used in the analyses differ in important and interesting ways. In general, the stress in the present paper is at the microeconomic level whereas the analysis in Weitzman (1985) is more aggregate in nature. There-

ture of the model reflects my interpretation of the fundamental aspects of a share system. This interpretation of Weitzman's argument is set forth in the next section of this paper. This paper is not an attempt to formalize each and every element of Weitzman's argument.

The structure of the model economy emphasizes the interactions across the sectors of a two-sector economy. Introducing share contracts into one sector changes the response of that sector to adverse shocks and alters the nature of the interaction between the sectors. In some cases, the presence of share contracts creates price fluctuations in place of the quantity fluctuations prevalent in economies with predetermined wages. Proposition 1 outlines conditions under which quantity fluctuations in one sector are reduced by the introduction of share contracts in the other sector.

The main result of the paper concerns conditions under which the introduction of share contracts in one sector of the economy increases the welfare of all agents. Proposition 2 of the paper shows that it is possible to construct a share contract that yields an allocation *weakly* Pareto dominating that is obtained under a wage system. This result is supportive of Weitzman's argument that the introduction of a share system can be beneficial. However, Proposition 2 indicates that the share contract which leads to this weakly Pareto superior outcome is very special. That is, there is only one share contract which can balance the gains and losses to the various groups of agents in the economy from the introduction of share contracts. This result contrasts with that reported in Weitzman (1983,1984) by placing strict conditions on the parameters of a share system.

The formal model is presented and analyzed in Section II and its welfare implications are displayed in Section III. Section IV contains some critical comments on the exercise performed in this paper and some ideas for future research in this area.

I. Overview and Interpretation of the Model

This section provides an intuitive discussion of the arguments which Weitzman advanced in favor of the share system. The starting point of the analysis is Weitzman's view that the predominant mode of labor transactions predetermines the worker's wage (in terms of the numeraire) and allows the firm to choose employment levels. Weitzman terms this structure a *wage system*. Even though the wage rate is influenced by competitive forces, *ex ante*, it is assumed to be inflexible *ex post*. Hence, any unforseen disturbances to the economy affect the level of employment but not wages. Through the circular flow of income, a multiplier effect arises which exacerbates the real effects of an initial shock. Weitzman (1984) contends that,

> The current wage system of compensating labor is a perilous anachronism that needs to be replaced. For when a contractionary impulse hits, not only is the initial response of a wage economy to throw people out of work, but a wage system can deepen a recession, multiplying its adverse consequences until the economy is trapped in a vicious cycle of persistent involuntary underutilization of the major factors of production. This public cost of the wage system—its "macroeconomic externality" of misbegotten unemployment spawning further unemployment —is a pollution-like consequence that private agents have little incentive to consider. [Weitzman, p. 46]

Based on this argument, Weitzman proposes the introduction of share contracts. In

---

fore, this paper provides more details on the interactions across sectors of the economy under alternative compensation schemes. This is particularly important given that this paper investigates the impact of share contracts in some but not all sectors of the economy whereas Weitzman (1985) considers share contracts in all sectors. Further, the welfare analysis in this paper is at the individual level rather than at the level of aggregate variables as in Weitzman (1985). This permits a more complete evaluation of the costs and benefits of a share system as well as a sharper characterization of the circumstances under which share contracts yield Pareto-superior outcomes. Other differences and similarities between the models will be brought out in the text when appropriate.

its simplest form, a share contract is characterized by two parameters: a constant transfer term ($\bar{\omega}$) and a share coefficient (λ). The contract also specifies some index of firm activity such as profits per worker, revenues per worker, etc., to which compensation is linked.[2] This index is assumed to be a *decreasing* function of the level of activity at the firm so that compensation per worker falls (assuming $\lambda > 0$) as the firm expands. Under either of these compensation systems, firms are given the latitude of choosing employment, *ex post*.

In a world of certainty, both of these compensation schemes can be shown to support the same labor allocation (see Proposition 1, Weitzman, 1983). However, there is a critical difference between the two systems. In a wage system, the labor market equilibrates at zero excess demand. In a share system, however, firms offering share contracts have an incentive to increase their work force further (at the equilibrium allocation) to take advantage of reductions in average wages associated with higher-employment levels. However, firms are prevented from expanding in this way by the existence of competing offers by other firms wishing to employ labor. Hence, as Weitzman puts it, share firms are in a state of excess demand for labor in the certainty equilibrium of the share economy (Section II elaborates on this point).

Because of this excess demand property, the share economy operates quite differently in the face of unanticipated fluctuations in demand. Since share firms are in a state of excess demand for labor, they are willing to employ those workers laid off by wage firms. This reduces fluctuations in employment and will eliminate the employment effects of the smaller shocks altogether. Weitzman (1984) argues that,

> A share system looks very much like a labor shortage economy. Share firms ever hungry for labor are always on the prowl—cruising around like vacuum cleaners on wheels, searching in nooks and crannies for extra workers to pull in at existing compensation parameter values. Such an economy inherently resists recession. [Weitzman, p. 98]

The upshot of this comparison is that a share economy is better able to absorb shocks without causing unemployment in labor markets. This, Weitzman argues, means that a share system socially dominates a wage system. Weitzman acknowledges that the wage system may privately dominate the share system for workers and firms so that a tax and transfer system will be necessary to induce those agents to trade labor services through socially superior share contracts.

In evaluating this argument, a number of issues come to mind. What is the macroeconomic externality that Weitzman refers to in his discussion of the wage system? Under what conditions and by what criteria will a share system dominate a wage system?

This paper addresses these points. The model outlined in Section II focuses on the interaction between agents in a multisector imperfectly competitive economy. This economy exhibits a "macroeconomic externality" in that actions taken in one sector of the economy directly influence the welfare of agents in other sectors. This economy also exhibits the multiplier property that Weitzman notes in his characterization of the wage system so that sector-specific shocks create aggregate fluctuations. Finally, the multisector setting allows one to evaluate the *external benefits* of introducing share contracts in a subset of the sectors. With regard to a welfare criterion, this paper focuses on the welfare of *individual* agents under alternative compensation schemes.

The main result in the paper is that the introduction of a share contract for *one* sector of the economy can support an allocation which weakly Pareto dominates that obtained under a wage system. A characterization of this share contract is provided in the paper. The key point is that by stabilizing output and employment in one sector, an appropriately designed share contract reduces (but does not necessarily eliminate) the spillover effects across sectors. As a consequence, aggregate fluctuations arising from sector-specific shocks are reduced. The cost of stabilizing output is that wages and prices fluctuate more in the presence of share contracts. Hence, workers in a share system face more risk in their compensation. Nonetheless, it is possible to design a share contract in which the benefits of quantity stabilization exceed these costs.

---

[2] Weitzman (1985) focuses on profit-sharing schemes while the analysis here is mainly devoted to revenue-sharing schemes. Weitzman (1983) discusses a broad range of alternative compensation schemes.

Other than attempting to explicitly model Weitzman's "macroeconomic externality," the model follows the assumptions in his earlier papers. In particular, the wage system is taken as a characterization of current labor market relations. Thus we do not provide a microeconomic rationale for the wage system. The potential problems with this approach are discussed in Section III. In addition, Weitzman views shocks to the economy as being totally unforeseen by agents. We comment on the importance of this restriction as well.

II. The Basic Model

The modeling approach taken here to compare wage and share systems draws on a series of recent papers by Oliver Hart, 1982; Weitzman, 1982, 1985; Walter Heller, 1986; D. John Roberts, 1984; and Russell Cooper and Andrew John, 1987. These models, in one way or another, stress the importance of imperfect competition in understanding certain key Keynesian features including multiple equilibria, inefficient allocations, and multiplier effects. Crucial to these model economies is the circular flow of income created by specifying particular patterns of production and consumption in a multisector economy.[3] In particular, producing agents are viewed as specialists in their production activity and generalists in consumption. These two ingredients—imperfect competition and a circular flow of income—lead to situations in which quantities fluctuate widely relative to prices and in which multiplier effects tend to propagate shocks at the microlevel. We specify a model of this type below to capture the "macroeconomic externality" implicit in Weitzman's discussion.

To simplify matters, we consider an economy with two sectors indexed by $i=1,2$ producing distinct commodities. In each sector, there are F firms indexed by $f=1,2,\ldots,F$. The number of firms in each sector is fixed exogenously. There are also $2N$ identical workers who can work in either of the two sectors. Assume there is a total of three commodities in the economy: the two produced goods and one nonproduced commodity. The nonproduced commodity could be viewed as a proxy for money but, given the static nature of the model, such an interpretation could be misleading.[4] The nonproduced good will serve as numeraire.

Each firm is endowed with $\overline{m}$ units of the nonproduced good and has, at its disposal, a linear technology converting labor into output on a one-for-one basis. Letting the subscript "fi" denote firm f of sector i, profits for this firm (π_{fi}) are given by

$$\pi_{fi}=(p_i-w)q_{fi}=(p_i-w)l_{fi}.$$

Here p_i and w are the sector output price and wage respectively, q_{fi} is the firm's output, and l_{fi} is the level of employment.

The firm (equivalently its owner) spends its total income of $\pi_{fi}+\overline{m}$ on two commodities: the nonproduced good (m) and the produced good from the *other* sector (c_{-i}). The firm's preferences are given by $U(c_{-1}^{\alpha}m^{1-\alpha})$, where $0<\alpha<1$ and $U(\cdot)$ is strictly increasing. Firm's demands for c_{-i}

---

[3] This multisector setting is common to models of coordination failures referenced above. The key feature of these models is that quantity decisions by firms in one sector depend on the quantity decisions of firms in the other sector. These interactions arise because goods are normal so that income generated in one sector is spent on another. A one-sector version of the model would miss these interactions and would not allow an evaluation of the effects of alternative compensation schemes in one sector on the other sector(s).

[4] The nonproduced good creates a leakage out of the circular flow of income so that a locally unique equilibrium will occur. Dynamic generalizations of this framework may provide the required environment to generate a role for savings and a demand for money to serve as this leakage.

and m are

$$c_{-i} = \frac{\alpha}{p_{-i}}(\pi_{fi} + \overline{m}_i),$$

$$m = (1-\alpha)(\pi_{fi} + \overline{m}_i).$$

Here, as stated above, the subscript in c_{-i} and p_{-i} refers to the sector in which the firm is a consumer. This structure of production and consumption highlights the circular flow of income and expenditures. For now we set $\overline{m}_i = \overline{m}$ for $i = 1,2$, so that the economy is symmetric.

Workers are endowed with a unit of leisure time that they supply inelastically to the firm. An employed worker in sector i receives compensation w and consumes some of the nonproduced good and some of the good produced in the *other* sector. This worker's preferences are represented by $V(c_{-i}^{\beta}m^{1-\beta})$ where $V(\cdot)$ is strictly increasing with $0 < \beta < 1$. The demand functions are

$$c_{-i} = \frac{\beta}{p_{-i}}w, \qquad m = (1-\beta)w.$$

An unemployed worker receives zero compensation and has utility of $V(0) = 0$.

With this structure of preferences, endowments, and technology, one can consider a wide variety of market structures. Fluctuations in endowments and preferences can then be introduced to investigate the comparative statics properties of the decentralized economy. The emphasis here is on the implications of alternative schemes for trading labor services on the form and magnitude of fluctuations stemming from shocks to the exogenous variables.

Given the symmetry in the problem, it is easy to compute the competitive equilibrium for this economy:

$$p_1^* = p_2^* = w^* = \frac{F}{N}\left(\frac{\alpha\overline{m}}{1-\beta}\right) \quad \text{and}$$

$$q_1^* = q_2^* = N/F.^5$$

In this equilibrium, wages and prices are equal across sectors and we have full employment. Variations in firms' endowments in one or both sectors create fluctuations in prices and in the distribution of workers across sectors. The economy remains at full employment.[6]

This competitive system has neither of the key ingredients in Weitzman's formulation: wage rigidities and imperfect competition. We next consider the consequences of imperfect competition under three alternative compensation systems: fixed wages, flexible wages, and share contracts.

A. *The Wage System*

To begin, we outline the imperfectly competitive equilibrium for an arbitrarily chosen wage, $\overline{w}$. The $2F$ firms play a noncooperative game using quantities of output (equivalently employment) as strategy variables. We then discuss the determination of $\overline{w}$ in this imperfectly competitive system and begin our study of the comparative static properties of the economy under alternative compensation schemes.

We focus on the decisions of firm f in sector i and use the symmetry of the agents to describe the symmetric Cournot-Nash equilibrium. This firm chooses its output level, q_{fi}, taking as given the output decisions of *all* other firms in the economy, to maximize π_{fi} subject to

$$p_i = \frac{R_{-i}}{q_{fi} + \sum_{k \neq f} q_{ki}}. \tag{1}$$

Here p_i is the prevailing price in sector i, where R_{-i} denotes the total amount of

---

[5] To see that this is an equilibrium, note that profits equal zero so that the total expenditure on good 1 equals $F\overline{m}\alpha + Nw^*\beta$. The total value of the supply of good 1 equals Np^*. These are equal when $p_1^* = w^* = F/N(\alpha\overline{m}/1-\beta)$. By symmetry, the market for good 2 clears. The market for $\overline{m}$ clears by Walras' law.

[6] Fluctuations in $\overline{m}_i$ will lead to contractions and expansions in the sectors since changes in endowments translate into movements in relative demand across sectors. If workers are perfectly mobile, the economy will remain at full employment.

numeraire to be spent by sector $-i$ agents on sector i output. From the preference structure described earlier,

$$R_{-i}=F\{\alpha\overline{m}_{-i}+q_{-i}(\alpha p_{-i}-\overline{w}(\alpha-\beta))\},$$

where q_{-i} is the level of output for each firm in sector $-i$. Note that the Cobb-Douglas preferences imply that the share of income being spent on sector $-i$ is *independent* of the price prevailing in that sector. This greatly simplifies the analysis (see Hart, 1982, as well on the role of this structure).

Assuming that it can hire as many workers as desired at $\overline{w}$, the firm chooses q_{fi} to satisfy

$$(2)\qquad (p_i-\overline{w})+q_{fi}\frac{\partial p_i}{\partial q_{fi}}=0.$$

To determine a symmetric Cournot-Nash equilibrium, solve (2) using (1) and the assumption that $q_{fi}=q_i$ for all f. The output level per firm in sector i is given by

$$(3)\qquad q_i=\frac{R_{-i}\eta}{\overline{w}F},$$

where $\eta\equiv(F-1)/F$ is a measure of competitiveness for this sector. Substituting (3) into (1) yields an equation for the sector i price, that is,

$$(4)\qquad p_i=\overline{w}/\eta.$$

Because of the linearity in costs and the constant elasticity of demand, in equilibrium prices are independent of the output levels in either of the sectors and are simply a markup over constant marginal cost, $\overline{w}$.

By symmetry (4) holds for sector $-i$ as well. Substituting it into the definition of R_{-i} and then into (3) yields a useful relationship between the level of activity per firm in sector i (q_i) for a given level of activity per firm in the other sector (q_{-i}),

$$(5)\qquad q_i=\overline{A}_{-i}+\gamma q_{-i}.$$

This "reaction curve" for sector i has a positive intercept of $\overline{A}_{-i}\equiv\eta\alpha\overline{m}_{-i}/\overline{w}$ and a slope of $\gamma\equiv(\alpha-\eta(\alpha-\beta))$.[7] Since α,η,β all lie between zero and one, $0<\gamma<1$.

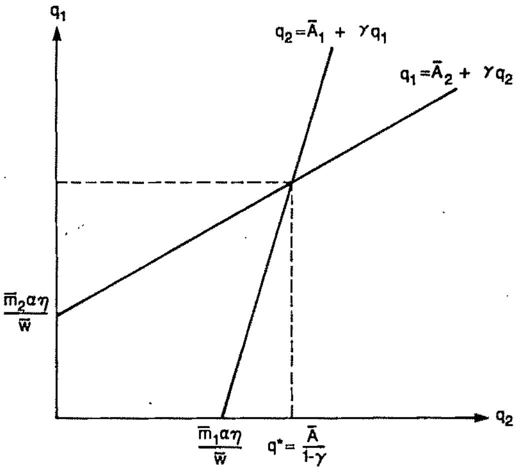

q_1
$q_2=\overline{A}_1+\gamma q_1$
$q_1=\overline{A}_2+\gamma q_2$
$\frac{\overline{m}_2\alpha\eta}{\overline{w}}$
$\frac{\overline{m}_1\alpha\eta}{\overline{w}}$
$q^*=\frac{\overline{A}}{1-\gamma}$
q_2

FIGURE 1

This curve is graphed in Figure 1. Again, invoking symmetry, there is an analogue to (5) giving q_{-i} as a function of q_i. This is shown in Figure 1 as well. The symmetric Nash equilibrium (q^*) is then given by the intersection of these curves,

$$(6)\qquad q^*=\frac{\overline{A}}{1-\gamma},\quad\text{with}\quad \overline{A}\equiv\frac{\eta\alpha\overline{m}}{\overline{w}}.$$

So, given $\overline{w}$, (6) characterizes the symmetric Nash equilibrium level of output for this economy.

There is a critical level of the wage, $w^c\equiv F\eta\overline{m}\alpha/N(1-\gamma)$, such that if $\overline{w}=w^c$ the equilibrium will be at full employment. This would be the case if there were an auctioneer controlling the real wage to ensure that the labor market cleared.[8] If $\overline{w}>w^c$, we have an underemployment equilibrium. For $\eta < 1$ (i.e., $F < \infty$), $w^c < w^*$. To keep the imperfect competition equilibrium at full employment requires a lower real wage as an inducement for the firms to increase output.

---

[7] This is not a proper reaction curve in that it incorporates the conditions for a Nash equilibrium in sector i given q_{-i}. Still, it indicates the reaction of sector i, in equilibrium, to changes in output of the other sector. We use the term "reaction curve" loosely.

[8] The fact that equilibrium in the labor market occurs at full employment even though firms are imperfect

With this structure in mind, we can begin our analysis of the comparative statics properties of this economy. In this analysis, the responsiveness of wages to disturbances in the economy will be crucial. To start, we follow Weitzman and assume that the wage is established to clear markets in the absence of shocks but is not responsive to unforeseen disturbances—this is the primary characteristic of the *wage system*. So, we set $\overline{w} = w^c$ to guarantee a long-run, full-employment equilibrium and then investigate the comparative statics for this model economy.

Suppose, for example, there is a decrease in the endowment of sector 2 firms, $\overline{m}_2$. This would cause the demand curve for sector 1 output to shift in, causing a reduction in output and employment in that sector. With sector 1 incomes reduced, the demand for sector 2 output will fall as well and a multiplier process ensues. As Figure 1 illustrates, the "reaction curves" of the firms are positively sloped and this feature is key in obtaining these multiplier effects.[9]

Starting at the full-employment, symmetric Nash equilibrium, the changes in the *equilibrium* levels of output are

$$\frac{dq_1}{d\overline{m}_2} = \frac{\frac{\eta}{\overline{w}}\alpha}{1-\gamma^2}; \tag{7}$$

$$\frac{dq_2}{dm_2} = \frac{\gamma\frac{\eta}{\overline{w}}\alpha}{1-\gamma^2} = \gamma\frac{dq_1}{d\overline{m}_2}.$$

The numerator in the $dq_1/d\overline{m}_2$ expression is the change in autonomous expenditures ($\overline{A}_2$) on sector 1 due to a reduction in $\overline{m}_2$. The denominator, $1-\gamma^2$, lies between zero and one and reflects the interaction between the two sectors. Output in sector 2 is reacting to sector 1 output reductions so that $dq_2/d\overline{m}_2 = \gamma dq_1/dm_2$. This process can be seen directly from Figure 1 by considering a shift in the q_1 "reaction curve." The multiplier for a change in autonomous expenditure is $1/1-\gamma^2$.

Hence employment fluctuations tend to feed on one another so that an initial shock to the system can create fairly large movements in employment. From the assumed rigidity of wages, unemployed workers wish to be working but the imperfectly competitive firms have no incentive to hire them given the downward-sloping demand curves they face.

The magnitude of this multiplier effect depends, ultimately, on γ. As α and β increase, more of an extra dollar of income (for either the firm *or* worker) is "returned" to the system so that leakages are reduced and the multiplier effect is increased. As the number of firms increases, η tends to 1 and γ tends to β. Thus the competitiveness of the economy influences the magnitude of the multiplier *if* there is a difference in consumption patterns between firms and workers (i.e., $\alpha \neq \beta$).

If wages were flexible in the short run, then $\overline{w}$ would adjust to equal the state-dependent value of w^c and therefore equilibrate the labor market given the firms' labor demand functions. In this case, the interaction between sectors would take the form of price rather than quantity movements as was the case in the competitive model with flexible wages described earlier. Thus the fact that Weitzman assumes the wage is fixed in terms of the numeraire is critical in this model since it leads to price inflexibilities and large output/employment fluctuations. This illustrates a point made by Hart (1982) that the underemployment and multiplier results in these models depend more on the presence of some imperfections in the labor market than on imperfectly competitive firms. Given the presence of fixed-real wage contracts, Weitzman's point about large

---

competitors occurs because labor is inelastically supplied.

[9] Hence the economy will exhibit behavior similar to those models with strategic complementarities investigated in Cooper and John (1987). The importance of the strategic complementarities becomes clear in the comparative static properties of the model.

movements in real quantities is substantiated in the model.

B. *The Share System*

We now consider a system with share contracts as an alternative to the wage system. Suppose that ρF of the firms in sector 1 employ workers under share contracts whereas all the other firms in sector 1 and all the sector 2 firms continue to trade with fixed-wage contracts. This structure allows us to evaluate the importance of the proportion of share firms on the magnitude of fluctuations in sectors 1 *and* 2 as we vary ρ between zero and one as suggested by Weitzman (1983). In contrast, Weitzman (1985) considers the impact of introducing share contracts in all sectors.

Weitzman's argument is that the presence of share firms will reduce employment fluctuations in sector 1 since these firms are always on the hunt for unemployed workers. Because of the "macroeconomic externality" across sectors of the economy, stabilization of output and employment in sector 1 may help stabilize sector 2 as well.

The analysis begins with a characterization of the equilibrium in an economy with share firms. This is done under the assumption that workers are perfectly mobile between wage and share firms, which is Weitzman's view of a long-run equilibrium. The results support those reported in Proposition 1 of Weitzman (1983) that the long-run equilibrium is independent of the structure of compensation. We then turn attention to the short-run response of the share economy to unforeseen shocks.

A share contract stipulates that wages depend inversely on some measures of firm activity. For simplicity, we take revenues per worker as that index which, given the firm's technology, is simply the price of output. Letting $w^s(p_1)$ be the wage paid to a worker at a share firm when the price is p_1,

$$w^s(p_1) = \tilde{w} + \lambda p_1, \quad 0 < \lambda < 1. \tag{8}$$

The share contract is characterized by a constant payment of $\tilde{w}$ and a share coefficient λ, which are *not* state dependent. Share contracts, though, do allow movements in the compensation level in response to price changes.

The experiment we are conducting is to investigate the effects on fluctuations in the economy of alternative (exogenously determined) contracts. Therefore think of the variables $(\rho, \tilde{w}, \lambda)$ as being imposed by an outside agency, such as a government. As discussed below, these parameters must satisfy certain restrictions in order for share firms to attract workers. The ultimate point of the exercise is to determine whether there exists an appropriately designed share system which can support a Pareto-dominating allocation.

To characterize the equilibrium, we have to consider the behavior of three classes of firms: share firms in sector 1, wage firms in sector 1, and wage firms in sector 2. A share firm takes the output decisions of all other firms as given and solves

$$\max_{q_1^s} \left(p_1(1-\lambda) - \tilde{w} \right) q_1^s,$$

where p_1 is the price that clears the sector 1 goods market. Note that $(\tilde{w}, \lambda)$ are taken as given by the share firm.

The first-order condition for the share firm's optimization problem is

$$p_1(1-\lambda) - \tilde{w} + q_1^s(1-\lambda)\frac{\partial p_1}{\partial q_1^s} = 0. \tag{9}$$

This condition is similar to (2) though the inclusion of $\lambda > 0$ is critical. As discussed in more detail below, the output level solving (9) may lead compensation (described by (8)) to fall beneath other opportunities for workers. So (9) holds if and only if share firms face an unlimited supply of workers. The inclusion of a labor supply constraint into the analysis will follow shortly.

Using (1), (9) can be rewritten as

$$\frac{R_2(1-\lambda)}{F(\rho q_1^s + (1-\rho) q_1^w)} \times \left(1 - \frac{q_1^s}{F(\rho q_1^s + (1-\rho) q_1^w)} \right) = \tilde{w}. \tag{10}$$

This equation implicitly defines the optimal quantities of the share firms as a function of the wage firm's output and the amount of spending by the other sector R_2.[10]

The reaction curves by the wage firms is given by

$$(11)\quad \frac{R_2}{F(\rho q_1^s+(1-\rho)q_1^w)} \times\left(1-\frac{q_1^w}{F(\rho q_1^s+(1-\rho)q_1^w)}\right)=\bar{w}.$$

This equation comes directly from (2). These two equations combined with (1) for $i=1$ completely describe quantities and the price in sector 1. The behavior of the sector 2 firms is again given by (3) and (4) with $i=2$.

To finish the statement of an equilibrium, we need to describe the behavior of workers in terms of their choice of employer. In particular, something needs to be said about workers' evaluation of consumption risks associated with employment by the different types of firms. At this stage, Weitzman (1984) argues that the shocks to the system are completely unpredictable and constitute uncertainty as opposed to risk. Hence agents completely ignore these shocks and evaluate the compensation schemes offered by alternative firms as if they lived in a certain environment. So, to describe the equilibrium, we need to determine the level of compensation that clears the labor market, given agents' beliefs that they face no risks.

Take, as a benchmark, the full-employment equilibrium with $\bar{w}=w^c$ and p_1 and $p_2=p^{**}$. Call the corresponding output level q^{**}, which is determined from (6) with $\bar{w}=w^c$. Furthermore, assume that $(\tilde{w},\lambda)$ satisfy

$$(12)\quad w^s(p^{**})=\tilde{w}+\lambda p^{**}=\bar{w}=w^c.$$

Hence, workers are indifferent between being employed by share or wage firms at the full-employment allocation.

So, if we follow Weitzman and assume agents ignore shocks, (12) is the appropriate labor market condition. If $(\tilde{w},\lambda)$ did not satisfy (12), then workers would be unwilling to accept positions at share firms.

In the event that Weitzman's view that shocks to the economy are entirely unpredictable is not correct, (12) is not the appropriate labor market equilibrium condition. Instead, workers at share firms would have to be compensated for the additional wage risk that they face and workers at wage firms would require compensation for employment risk. An alternative interpretation of (12) is that the outside agency which imposes the share contracts on ρF of the sector 1 firms also sets $(\tilde{w},\lambda)$ so that (12) holds and directly allocates workers to firms. Again, the interest here is whether there exists a share contract which yields a Pareto-superior allocation.

The full-employment allocation will be an equilibrium for all combinations of $(\tilde{w},\lambda)$ satisfying (12). This is Proposition 1 in Weitzman (1983).

To see this, suppose that $q_1^w=q_2=q^{**}$ and $p_1=p_2=p^{**}$. This will be an equilibrium if share firms choose to set $q_1^s=q^{**}$ as well, since $q_1^w=q_2=q^{**}$ is a best response for each of the other firms in the economy. Inserting (12) into (9) and evaluating this at $q_1^s=q^{**}$ yields

$$(13)\quad (p^{**}-\bar{w})+q^{**}(1-\lambda)\frac{\partial p_1}{\partial q_1^s}.$$

Since, by construction, q^{**} solves (2) at $p_i=p^{**}$, (13) cannot equal zero if $\lambda>0$. Share firms wish, in fact, to expand output beyond q^{**} since they share the loss of a lower price with their workers. However, if one of them did, p_1 would fall below p^{**} and $w^s(p_1)$ would fall below $\bar{w}$. Share firms would then be unable to attract any workers given the competing offers of wage firms at $\bar{w}$.

This point is shown in Figure 2. Given $\bar{w}$, wage firms select q^{**} as their desired level of output and employment. A share contract satisfying (12) is depicted as well. The desired level of output and employment by a share firm is given by $\tilde{q}_1^s>q^{**}$, where $\tilde{q}_1^s$

---

[10] We are assuming here that all wage firms and share firms produce the same levels of output: q_1^w and q_1^s, respectively.

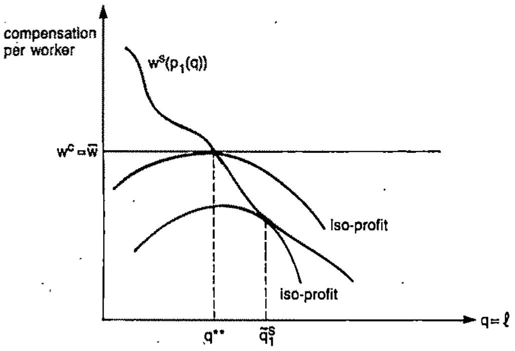


FIGURE 2

solves (9). Share firms are limited to producing only q^{**} by (12). (Note that the curves for the iso-profits are drawn for given levels of output by all other firms.)

Hence q^{**} is an equilibrium in which share firms find that (12) acts as a binding constraint on their output/employment choice. In this equilibrium, share firms are in a state of excess demand for workers.

With this full-employment equilibrium as a benchmark, we can repeat our earlier experiment and consider the response of this economy to a reduction in $\overline{m}_2$. Initially, the wage firms will reduce their employment and output. Share firms will hire the unemployed workers to take advantage of the reduction in compensation from an increase in output. For small enough reductions in $\overline{m}_2$, there will be no unemployment in sector 1. Instead prices will fall in response to the decrease in $\overline{m}_2$. This price reduction *will* influence the demand curve facing sector 2. Hence, the externalities associated with contractions in one sector leading to contractions in the other sector are not avoided in the share economy. A share system changes the nature of these spillovers but does *not* eliminate them. Nonetheless, a share system will be more stable for certain parameter configurations.

To formalize this intuition, we need to characterize the Nash equilibrium after the realization of a low $\overline{m}_2$. Let $\tilde{q}_1^s(q_1^w, R_2)$ solve (9) with equality. This is the level of output share firms would produce given (q_1^w, R_2) *if* they faced an unlimited supply of workers. The actual level of share firms output is given by

$$(14)\quad \hat{q}_1^s(q_1^w, R_2) = \min\left(\tilde{q}_1^s(q_1^w, R_2), \frac{N-(1-\rho)q_1^w F}{\rho F}\right).$$

This expression is the short-run reaction curve for share firms. The first term on the right side is the desired level of output for these firms from (9). The second term is the number of workers per share firm available in sector 1 *after* the wage firms have determined their employment level.[11] At the full-employment equilibrium (our benchmark) the second term is less than the first. This will continue to be the case for small enough reductions in $\overline{m}_2$ below $\overline{m}$. Hence, the share firms employ all of the workers in the sector *not* employed by the wage firms. These workers will accept jobs at the share firms even though they are paid less than the employed workers at the wage firms. There are *no* vacancies at the wage firms and the wage level is (by assumption) fixed. So, together, (11) and (14) describe the distribution of workers between share and wage firms for given levels of R_2.

For small enough fluctuations in $\overline{m}_2$, the reaction of sector 1 will be solely in terms of prices since, by (14), share firms ensure that q^{**} is the average level of output in the sector. Sector 1 price, p_1, is determined from

$$(15)\qquad p_1 = \frac{R_2}{Fq^{**}}.$$

The behavior of sector 2 firms is still characterized by (3) and (4), and this determines R_2. Differentiating the conditions for equilibrium with respect to $\overline{m}_2$, we can derive that (the superscripts indicate that these conditions hold for the share system):

---

[11] Unemployed workers in sector 2 are not able to work for sector 1 share firms in the short run. This restriction reflects the short-run immobility of workers due to sector-specific training. Allowing sector 2 workers to be employed by sector 1 share firms would further strengthen the vacuum cleaner effects reported here.

$$(16)\qquad \Delta q_2^s = \frac{\eta}{\bar{w}F}\Delta R_1^s,$$

$$(17)\qquad \Delta R_1^s = F\phi q^{**}\Delta p_1^s,$$

$$(18)\qquad \Delta p_1^s = \frac{1}{q^{**}}\left(\frac{\bar{w}}{\eta}\gamma\Delta q_2^s + \alpha\Delta\bar{m}_2\right).$$

Equation (16) is the quantity-response function in sector 2 to a change in expenditures on that sector and is derived directly from (3). Equation (17) relates the change in expenditures to the change in the sector 1 price. The variable $\phi \equiv \alpha - \rho\lambda(\alpha-\beta)$ and lies between zero and one.[12] Finally, (18) describes a change in the sector 1 price due to the exogenous change in sector 2 endowments and the endogenous change in sector 2 output and is derived from (15). Recall that $\gamma \equiv \alpha - \eta(\alpha-\beta)$. Since $p_2 = \bar{w}/\eta$ and $\rho q_1^s + (1-\rho)q_1^w = q^{**}$, this completes the description of the comparative statics effects of a *reduction* in $\bar{m}_2$.

Substituting (18) into (17) and (17) into (16) yields

$$(19)\qquad \frac{\Delta q_2^s}{\Delta\bar{m}_2} = \frac{\eta\phi\alpha/\bar{w}}{1-\gamma\phi} \quad \text{and}$$

$$(20)\qquad \frac{\Delta p_1^s}{\Delta\bar{m}_2} = \frac{\alpha/q^{**}}{1-\gamma\phi}.$$

Equations (19) and (20) are the counterparts of (7) for the share economy. In comparing (19) and (20) with (7), we note that fluctuations in the share system take the form of both price and quantity movements. The share sector responds to shocks through relative price changes and these produce movements in sector 2's output and employment level. To stress an important point, share contracts do not eliminate spillover effects across sectors of the economy. Instead, they alter the form of these interactions. With these comparative static results in mind, we now turn to a welfare evaluation of the share system.

III. Welfare Properties of the Share System

A full welfare evaluation of the share system should start with a complete statement of the primitives for our economic model. One could then ideally proceed to a comparison of the decentralized allocation with the set of Pareto-optimal allocations.

Our evaluation of the share system falls short of this methodology in a number of important respects. First, our comparison is between the wage and share systems. We have not argued that the wage system is a privately optimal contracting arrangement for this economy. We will not assert that a share system can support a Pareto-optimal allocation. Instead, the argument for this comparison often given is that the wage system is empirically relevant and possibly responsible for magnifying shocks to the economy. Consequently, it is useful to see how that system compares with alternatives such as the share system.

This approach is potentially dangerous in that the ultimate argument to support the wage system (assuming it is empirically relevant) may impose constraints which make the share system more costly to implement.[13] Without a theory predicting the wage system, it is impossible to gauge the importance of this concern. We nonetheless proceed with this comparison and discuss some approaches to understanding the wage system in the conclusion.

A second issue concerns the treatment of uncertainty and risk sharing in the model. In specifying preference, both workers and firms were allowed to be risk averse. The actual economic behavior described so far has avoided any discussion of risk sharing. This is consistent with Weitzman's view that the randomness affecting the macroeconomy is totally and completely unpredictable so that agents behave as if they lived in a world of certainty. We continue to ignore these risk-sharing considerations and assume that both $U(\cdot)$ and $V(\cdot)$ are linear.

---

[12] To derive (17), use the definition of R_1 given earlier to derive R_1^s which is the amount of numeraire spot by sector 1 agents on sector 2 output in a share system. Total differentiation of R_1^s with respect to p_1^s, using $dw^s(p_1)/dp_1 = \lambda$ will yield (17).

[13] For example, contracting models with asymmetric information predict that firms should be given the latitude of choosing employment *ex post* as in the wage system. This asymmetry of information could also make the share system more difficult to implement.

Without completely characterizing the planning problem and without a complete treatment of uncertainty, it is still possible to investigate some of the welfare properties of the two systems. Our approach will be to look at changes in utility levels for firms and workers due to reductions in $\bar{m}_2$ *across* the two systems.

To begin, we consider the effects of changing from a wage system to a share system in sector 1 on the magnitude of fluctuations in sector 2. As noted earlier, in selecting compensation schemes and employment rules, agents will not take into account the effects of their contracting structures on the rest of the economy. Hence it is important to see whether there are any external benefits to sector 2 agents from having share contracts in sector 1. To do so, we will continue to assume that labor in sector 2 is traded under a wage system.

Letting superscripts denote the system (s for share, w for wage), we can compare the equilibrium changes in *sector 2* output and employment across the two systems from

$$(21)\quad \Delta \equiv \frac{\Delta q_2^w}{\Delta \bar{m}_2} - \frac{\Delta q_2^s}{\Delta \bar{m}_2} = \frac{n\alpha}{\bar{w}}\left[\frac{\gamma}{1-\gamma^2} - \frac{\phi}{1-\gamma\phi}\right].$$

Equation (21) implies that $\gamma > \phi$ is necessary and sufficient for output in sector 2 to be more stable in a share economy than in a wage system. From the definitions of γ and ϕ, $\gamma > \phi$ means that

$$(22)\quad \eta(\alpha-\beta) < \rho\lambda(\alpha-\beta).$$

PROPOSITION 1: *If $\alpha > \beta$, share contracts help stabilize sector 2 output as long as $\rho\lambda > \eta$. If $\alpha \leq \beta$, share contracts are destabilizing.*

PROOF:

Direct from (21) and (22).

Thus one role of share contracts, in this model, is to exact a redistribution of income from workers to firms in economic downturns. If $\alpha > \beta$, firms tend to "recycle" more income than workers so that the leakage effect in downturns is reduced under the share system. So one interpretation of this is that share contracts act as an automatic stabilizer.

The workers in sector 1 also appear to be better off under a share system in the sense that their employment is stabilized. This is true even if the conditions for Proposition 1 do not hold. Hence, if we focus exclusively on fluctuations in employment and output, the share system is welfare improving for workers in sector 1. Nonetheless, when $\alpha < \beta$, the introduction of share contracts in sector 1 may create larger fluctuations in sector 2 employment levels. Thus, in contrast to Weitzman (1985), share contracts in one sector could exacerbate fluctuations in another sector.

This concentration on quantity fluctuations, however, misses some of the costs of the share system. From a worker's viewpoint, at a share firm, the price fluctuations induced by the share system imply that wages fall in downturns. It is not obvious that workers prefer the employment security of share contracts to the wage security and employment risk of the wage system.[14] To better understand this tradeoff between employment compensation and price risk, we need to evaluate the utility changes of the agents in question.

Looking first at workers, we must consider the changes in utility levels from changes in $\bar{m}_2$ across the two systems for three types of workers: sector 1 workers at wage firms, sector 1 workers at share firms, and sector 2 workers. A sector 1 share worker is always *worse* off in low $\overline{m}_2$ states than a sector 1 worker at a wage firm since the latter have a chance of a job at $\overline{w}$ before becoming employed at a share firm, whereas the former always face reductions in compensation when $\overline{m}_2$ falls. Hence, if we can show that share workers in a share system and sector 2 workers do better under a share system, then we can conclude that all workers do better in a share system.

---

[14] Weitzman (1985) concentrates on aggregate-quantity fluctuations but does not discuss wage variations. Thus some of the welfare effects of introducing share contracts are missed in that analysis.

A. *Sector* 1 *Workers*

The change in the utility of a sector 1 worker in a *wage system* due to a change in $\overline{m}_2$ is given by

$$(23) \qquad \frac{\Delta V_1^w}{\Delta \overline{m}_2} = \frac{\mu \eta F}{N p_2^{\beta}}\left[\alpha + \frac{\overline{w}\gamma}{\eta}\,\frac{\Delta q_2^w}{\Delta \overline{m}_2}\right],$$

where $\Delta q_2^w/\Delta \overline{m}_2$ is the equilibrium change in sector 2 output from decreases in $\overline{m}_2$ in a *wage* system (see (7)) and $\mu \equiv \beta^{\beta}(1-\beta)^{1-\beta}$. Equation (23) utilizes the assumption that each worker faces an equal chance of being laid off under a wage system.[15]

The change in the utility of a sector 1 worker for a share firm in a share system is given by

$$(24) \qquad \frac{\Delta V_1^s}{\Delta \overline{m}_2} = \frac{\mu \lambda F}{N p_2^{\beta}}\left(\alpha + \frac{\overline{w}\gamma}{\eta}\,\frac{\Delta q_2^s}{\Delta \overline{m}_2}\right),$$

where $\Delta q_2^s/\Delta \overline{m}_2$ is given by (19). Denote by $\hat{\Delta}$ the difference between (23) and (24).

Comparing (23) and (24), workers under either system benefit if the equilibrium output responses to a reduction in $\overline{m}_2$ are small.

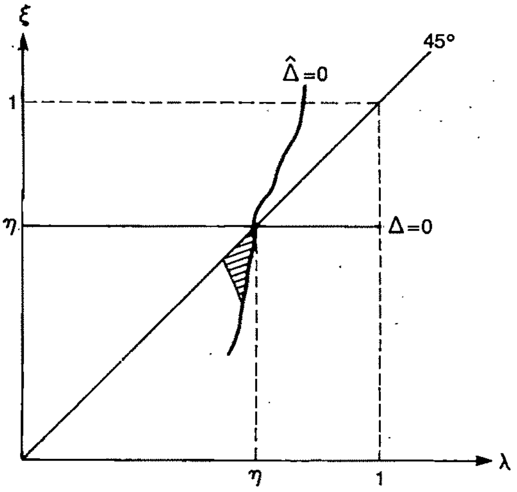


FIGURE 3

For the calculation in (23), low levels of $\Delta q_2^w/\Delta \overline{m}_2$ help stabilize sector 1 employment. From (18), we see that changes in p_1 due to changes in $\overline{m}_2$ are smaller if $\Delta q_2^s/\Delta \overline{m}_2$ is smaller. Share workers gain from stabilizing $\Delta q_2^s/\Delta \overline{m}_2$ since price reductions lower compensation if $\lambda > 0$. For a given level of $\Delta q_2^s/\Delta \overline{m}_2$, share workers prefer lower values of λ. If $\lambda = \eta$ and $\rho = 1$, then from (21), we know that (23) and (24) are equal and $\hat{\Delta} = 0$.

It is useful to think of policymakers having control over two variables ρ and λ, with $\overline{w}$ adjusting so that (12) holds. Given the importance of $\rho\lambda$ as a product in ϕ and hence in (19) and (22), consider $\xi \equiv \rho\lambda$ and λ as the policy variables. The horizontal line in Figure 3 at $\xi = \eta$ guarantees that $\Delta q_2^w/\Delta \overline{m}_2 = \Delta q_2^s/\Delta \overline{m}_2$, or $\Delta = 0$ from (21). Note too that since $\rho \le 1$, $\xi \le \lambda$, we are restricted to combinations of λ and ξ on or below the 45° line. The line labeled $\hat{\Delta} = 0$ contains the combination of (λ, ξ) such that (23) and (24) are equal. This curve goes through the point (η, η) as shown in Figure 3. From (23) and (24), it can be shown that the $\hat{\Delta}$ curve has a slope exceeding one in the neighborhood of (η, η). This is depicted in Figure 3 as well. Above the $\hat{\Delta} = 0$ curve, we find that $\Delta V_1^w/\Delta \overline{m}_2 > \Delta V_1^s/\Delta \overline{m}_2$ so that the share system provides better protection for sector 1 workers against adverse $\overline{m}_2$ shocks.

---

[15] The probability that an arbitrary sector 1 worker is employed is given by Fq_1/N, where q_1 is the average employment level. Expression (23) is obtained by differentiating a worker's expected utility with respect to $\overline{m}_2$ where the uncertainty comes from a worker's employment status. The model ignores the role of seniority in layoff decisions. As suggested by one of the referees, more senior workers generally face little layoff risk and hence would be less likely to support the movement to a share system.

B. *Sector* 2 *workers*

For sector 2 workers, we calculate that

$$(25)\quad \frac{\Delta V_2^w}{\Delta \overline{m}_2} = \frac{\mu F}{p_1^\beta N}\overline{w}\frac{\Delta q_2^w}{\Delta \overline{m}_2}, \quad \text{and}$$

$$(26)\quad \frac{\Delta V_2^s}{\Delta \overline{m}_2} = \frac{\mu F}{p_1^\beta N}\overline{w}\frac{\Delta q_2^s}{\Delta \overline{m}_2} - \frac{\mu \overline{w}\beta}{p_1^{\beta+1}}\frac{\Delta p_1}{\Delta \overline{m}_2},$$

where ΔV_2^i is the change in sector 2 workers' utility under sector 1's use of compensation scheme *i*. As in the case of (23), (25) measures utility changes of sector 2 workers from the employment risk they face from variations in q_2^w. In (26), sector 2 workers are seen to face employment risks *and* variations in the price of their consumption good. One important difference between (25) and (26) is that sector 2 workers benefit from sector 1 price reductions in a share system when $\overline{m}_2$ falls (i.e., $\Delta p_1/\Delta \overline{m}_2 > 0$). Hence, even if employment fluctuations in the two systems are identical (so $\Delta q_2^w/\Delta \overline{m}_2 = \Delta q_2^s/\Delta \overline{m}_2$ from (21)), sector 2 workers still prefer a share system in sector 1.

The gains or losses for sector 2 workers from a share system can be seen from Figure 3 as well. Along the line $\Delta = 0$, we know that (25) exceeds (26), since $\Delta p_1/\Delta \overline{m}_2 > 0$. So, $\Delta V_2^w/\Delta \overline{m}_2 > \Delta V_2^s/\Delta \overline{m}_2$ for combinations of (λ, ξ) slightly below the line (by continuity).

Looking at Figure 3, any combination of (λ, ξ) in the shaded region (i.e., close to (η, η) with $\xi \le \lambda$) will determine a share contract in which workers in both sectors are better off than in the wage system. Since we are below the $\Delta = 0$ line, fluctuations in sector 2 output are actually larger in the share system than in the wage system (if $\alpha > \beta$). Sector 2 workers are compensated for this by reductions in sector 1 output prices when $\overline{m}_2$ is low. Sector 1 workers have their employment stabilized and their compensation destabilized. Above the $\hat{\Delta} = 0$ line, λ is low enough, relative to ξ, that sector 1 workers are better off with slightly destabilized compensation. Workers, then, are in agreement about the merits of stabilizing the system through share contracts, that is, both groups of workers benefit from large values of ξ. Sector 1 workers, of course, prefer that this be accomplished by making all firms share firms (i.e., $\rho = 1$) rather than through large λ. Given this, we let $\rho = 1$ for the remainder of the analysis since our ultimate goal is to show there exists a share system preferable to the wage system.

C. *Sector* 1 *Firms*

With regard to firms, we first look at those in sector 1. Under a wage system, we can calculate that

$$(27)\quad \frac{\Delta U_1^w}{\Delta \overline{m}_2} = \frac{\Psi}{p_2^\alpha}(1-\eta) \times \left(\alpha + \frac{\overline{w}\gamma}{\eta}\frac{\Delta q_2^w}{\Delta \overline{m}_2}\right),$$

where $\Psi \equiv \alpha^\alpha(1-\alpha)^{1-\alpha}$. In a share system, we use (18) to calculate that

$$(28)\quad \frac{\Delta U_1^s}{\Delta \overline{m}_2} = \frac{\Psi}{p_2^\alpha}(1-\lambda) \times \left(\alpha + \frac{\overline{w}\gamma}{\eta}\frac{\Delta q_2^s}{\Delta \overline{m}_2}\right).$$

So, like their workers, sector 1 firms benefit from stabilizing fluctuations in sector 2 output. These firms, however, prefer that λ be large so that they do not have to absorb large profit reductions in bad states.

Define $\tilde{\Delta}$ as the difference between (27) and (28). The combinations of (λ, ξ) such that $\tilde{\Delta} = 0$ are shown in Figure 4. Note that this curve goes through the point (η, η) as well. In contrast to $\hat{\Delta}$, $\tilde{\Delta} = 0$ is negatively sloped, which reflects firms' preference for higher λ, given ξ. Comparing $\tilde{\Delta} = 0$ to $\hat{\Delta} = 0$, we see that there are no combinations of (λ, ξ) with $\xi \le \lambda$ such that *all* sector 1 agents are strictly better off with share contracts. At the point (η, η) both sector 1 workers and firms are indifferent between the two systems.

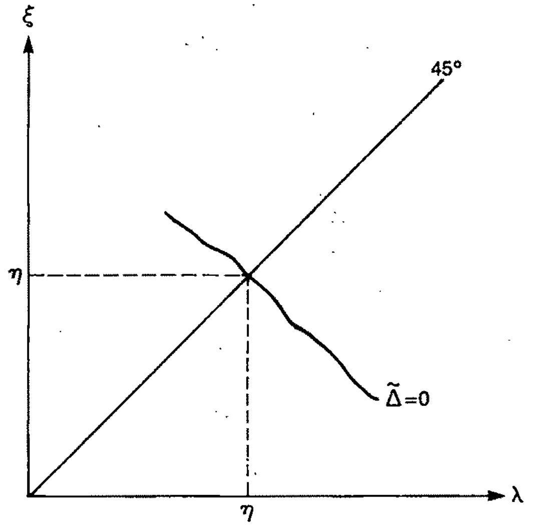


FIGURE 4

D. *Sector 2 Firms*

Finally, the utility changes for sector 2 firms due to reductions in $\overline{m}_2$ are given by

$$(29)\quad \frac{\Delta U_2^w}{\Delta \overline{m}_2}=\frac{\Psi}{(p_1)^\alpha}\times\left(1+\frac{(1-\eta)\overline{w}}{\eta}\frac{\Delta q_2^w}{\Delta \overline{m}_2}\right),$$

and

$$(30)\quad \frac{\Delta U_2^s}{\Delta \overline{m}_2}=\frac{\Psi}{(p_1)^\alpha}\times\left(1+\frac{(1-\eta)\overline{w}}{\eta}\frac{\Delta q_2^s}{\Delta \overline{m}_2}-(\pi_{f2}+\overline{m}_2)\frac{1}{p_1}\frac{\Delta p_1}{\Delta \overline{m}_2}\right).$$

As was the case for sector 2 workers, these firms gain from stabilizing their output and from reductions in p_1 when $\overline{m}_2$ is low. For (λ,ξ) sufficiently close to (η,η), (29) will exceed (30).

E. *Welfare Results*

These calculations make quite clear that all agents in the economy benefit from the stabilization of output in both sectors (assuming that Proposition 1 holds) associated with a share system. These are the gains to the share system stressed in Weitzman (1985). The potential losers from the share system are sector 1 workers who bear the burden of compensation reductions in bad states. Is it possible to devise a share contract such that the share system dominates the wage system?

PROPOSITION 2: *Agents' utility losses when $\overline{m}_2$ falls are less under a share system with $\rho=1$ and $\lambda=\eta$ than under a wage system.*

PROOF:

With $\rho=1$ and $\lambda=\eta$, $\xi=\eta$ as well. Hence, from (21), $\Delta=0$ (regardless of the sign of $\alpha-\beta$). At $\lambda=\eta$ and $\rho=1$, all sector 1 agents suffer identical utility losses, when $\overline{m}_2$ is low, in the two systems. At $\lambda=\eta$ and $\rho=1$, sector 2 agents are better off with the share system since $\Delta=0$ and $\Delta p_1/\Delta\overline{m}_2>0$.

In thinking about this proposition, note that at $\lambda=\eta$ and $\rho=1$, fluctuations in sector 2 output are identical under the two systems. All agents in sector 1 are indifferent between the two compensation systems: loosely speaking, the stabilization in employment and output brought about by share contracts is just offset by variations in prices and wages they generate. By construction, the gains from the share system come from the increased consumption of sector 2 agents when sector 1 output and employment are stabilized.

The proposition tells us about the welfare properties of the two systems when $\overline{m}_2$ is lower than $\overline{m}$. If $\overline{m}_2$ is unexpectedly high, then the two systems will generate the same outcome since, at the benchmark allocation described by (6) with $\overline{w}=w^c$ and $\overline{m}_1=\overline{m}_2=\overline{m}$, we are at full employment. That is, an increase in $\overline{m}_i$ will simply raise prices in the economy while output remains at full employment. Hence by setting $\rho=1$ and $\lambda=\eta$, the share system generates an allocation that stochastically dominates (first-order) the allocation supported by the wage system. So that, even if agents were strictly risk averse, the share system would be better than the wage system.

An important implication of Proposition 2 is that if $\rho = 1$ but $\lambda \neq \eta$, then the share system will not lead to a Pareto-improving allocation. Thus to reap the benefits of introducing share contracts requires rather exact information on the competitiveness of the economy. This contrasts with the results reported by Weitzman (1983, 1985) in which the share contract leading to better outcomes is not restricted.

Finally, note that the degree of imperfect competition, measured by η, influences the sensitivity of wages to prices in the share contract. So as the economy becomes more competitive, $\eta \to 1$ and share workers have more variation in their compensation.

IV. Concluding Comments

The main point of this exercise has been to establish a framework for evaluating Weitzman's proposal for the imposition of a share system. The model highlights multiplier effects in quantity responses under a wage system. Following the arguments in Cooper and John (1987), these multiplier effects were driven by strategic complementarities across sectors of the economy. From this perspective, the value of share contracts is their potential for altering the form and magnitude of intersectoral linkages. In this model, share contracts in one sector may have positive external benefits for other sectors.

The main result of the analysis is contained in Proposition 2, which shows that with information about the degree of competitiveness in the economy, it is possible to construct a *particular* share contract that yields an allocation Pareto dominating the equilibrium in a wage system, but that other share contracts do not improve welfare. Thus the results here contrast with the view expressed by Weitzman (1983–1985) on the benefits of a share system of any kind.

A few words of caution are perhaps in order. First, the model used in this paper is quite special in a number of respects: demands were iso-elastic and linear in income and the technology was linear as well. Whether the stabilization of quantities brought about by share contracts will *generally* outweigh the cost to workers in terms of the extra risks they face remains to be demonstrated. Second, the model excludes any discussion of the impediments to the share system—see Weitzman (1984) for a discussion of these potential costs.

The ultimate value of this exercise depends on whether or not currently utilized compensation systems are responsible for apparent macroeconomic inefficiencies. While the stress in this paper and in the papers by Weitzman has been on the problems of the wage system, the more general question concerns the social optimality of privately optimal contracts. This paper has highlighted a "macroeconomic externality" across sectors of an imperfectly competitive economy. In the presence of such an externality, privately optimal arrangements for the trade of labor services will not necessarily reflect the public good, and interventions may be appropriate.

This paper should thus be viewed as an initial attempt to understanding the nature of this "macroeconomic externality" and its effects on the social optimality of certain labor arrangements. It is certainly possible to argue that the comparison made in this paper between the wage and share system is really irrelevant because of a lack of theoretical and empirical support for the wage system. An alternative view is that the basic macroeconomic properties of the wage system hold even if firms and their workers sign optimal labor contracts so that the quest for Pareto-superior compensation schemes is a useful and important exercise. This paper is then a first step toward addressing the more general problem of analyzing the social optimality of privately optimal labor contracts.

REFERENCES

Cooper, Russell and John, Andrew, "Coordinating Coordination Failures in Keynesian Models," Cowles Foundation Discussion Paper No. 745R, Yale University, revised

1987, *Quarterly Journal of Economics*, forthcoming 1988.

Hart, Oliver, "A Model of Imperfect Competition with Keynesian Features," *Quarterly Journal of Economics*, February 1982, *97*, 109–38.

Heller, Walter, "Coordination Failure Under Complete Markets with Applications to Effective Demand," in *Equilibrium Analysis: Essays in Honor of Kenneth J. Arrow*, Vol. II, Walter P. Heller, Ross M. Starr, and David A. Starrett, eds., Cambridge: Cambridge University Press, 1986.

John, Andrew and Nordhaus, William, eds., "The Share Economy: A Symposium," *Journal of Comparative Economics*, December 1986, *10*, 414–73.

Roberts, D. John, "General Equilibrium Analysis of Imperfect Competition: An Illustrative Example," mimeo., Stanford University, 1984.

Weitzman, Martin, "Increasing Returns and the Foundations of Unemployment Theory," *Economic Journal*, December 1982, *92*, 787–804.

______, "Some Macroeconomic Implications of Alternative Compensation Systems," *Economic Journal*, December 1983, *93*, 763–83.

______, *The Share Economy*, Cambridge, MA: Harvard University Press, 1984.

______, "The Simple Macroeconomics of Profit Sharing," *American Economic Review*, December 1985, *75*, 937–53.

Editorial, "Best Idea Since Keynes," *New York Times*, 28 March 1985.

The Economics of Rising Stars

By GLENN M. MACDONALD*

The paper provides analysis of an occupation in which there is substantial uncertainty about individual performance, and past performance is correlated with future outcomes. In equilibrium, only the young enter the occupation and only the successful stay on. The young are numerous and earn incomes well below what their current alternative offers; success is rare and rewarded highly. The implied distribution of earnings must have positive skewness.

Economic success, or lack thereof, is a central feature of every working individual's labor market experience. That the extent to which success is achieved—as measured by income, control over other inputs, stability of employment, and so on—is highly variable even within narrowly defined groups is a basic fact of empirical labor economics. This observation has also proved to be a very difficult one to explain adequately within a model in which income, for example, is the value of productive traits at parametric prices. Indeed, the distribution of rewards appears to be such that comparatively minor talent differences can generate enormous returns, so that the distribution of income is not a simple rescaling of the distribution of ability. Examples involving professional athletes and musicians come to mind immediately, but the familiar large variation and positive skewness in the distribution of earnings within less exotic occupations suggests that the same sort of mechanism is at work much more generally.

Sherwin Rosen (1981) demonstrated that equilibrium returns can greatly exaggerate talent differences. In his Superstars model, consumers value a unit of output in proportion to its producer's talent. For example, consumers of music prefer any given piece to be performed by a more able artist. The equilibrium unit price that a producer obtains is thus proportional to talent. As a consequence, because the price of other inputs does not depend on talent, abler producers choose to produce more units of output. Their net payoff is therefore an increasing, and more importantly, convex function of talent. The price of talent implicit in the market value of output—net returns per unit of talent—is an increasing function of talent, and causes the distribution of returns to be right-skewed relative to the distribution of ability.

The variable implicit pricing of attributes such as talent, and the associated nonlinear returns, are the distinguishing characteristics of "hedonic" models (of which the Superstars model is one) more generally. In these settings, it follows that the distribution of overall returns may have features quite different from the distribution of attributes. This result has proven useful in a wide variety of applications. See Rosen (1974, 1987), and for more discussion, James Heckman and Jose Scheinkman (1987). At the same time, despite their generality and usefulness in organizing the data, existing hedonic models share a characteristic that limits their predictive content: the distribution of underlying traits is quite unrestricted. To illustrate, as explained above, the Superstars model yields earnings that are increasing and convex in talent, implying that the distribution of earnings has positive skewness in comparison to that of talent. But because the distribution of talent is effectively arbitrary, while the model is consistent with a positively skewed distribution of earnings, it does not imply it.

---

*Professor of Economics, Centre for Decision Sciences and Econometrics, Social Science Centre, University of Western Ontario, London, Ontario, Canada, N6A 5C2. Comments from Gary Becker, Lorne Carmichael, Jim Davies, James Heckman, George Neumann, Michael Parkin, Chris Robinson, Sherwin Rosen, Alan Slivinski, and two anonymous referees were most helpful, as was discussion in the Chicago applications workshop.

The general goal of this paper is to demonstrate how the predictive content of hedonic environments may be augmented by placing further structure on the distribution of underlying traits. The vehicle through which this demonstration is carried out is a simple stochastic dynamic Superstars model, which is of independent interest as a model of occupational choice.

The model restricts the distribution of (implicitly) priced attributes by deriving it from an information accumulation process formally very similar to that studied by myself (1982), and much in the spirit of the work of William Johnson (1978) and Boyan Jovanovic (1982). Because the analogy with the entertainment setting is a helpful one, it will be utilized in what follows. Thus the producer is a "performer," consumers the "audience," and so on. However, it should be emphasized that much broader interpretations are available.

Proceeding in this fashion then, performance quality may be either high or low in each of two periods, where the outcome is stochastic, positively correlated over time for a given performer, and observed by all who are interested. The market generates a price for tickets to see new performers, as well as a ticket price for performers who have a track record. Taking prices as given, performers decide when, if at all, to perform, and what to do in response to first-period reviews. Audience members choose which, if any, performers to see, given the reviews (for those performers who have performed previously) and ticket prices.

In the model's steady-state equilibrium, performers only enter the industry when young, and stay only if they get good reviews. Net returns more than compensate for differences in information across performers, so that those who receive good reviews over time serve vast audiences and experience dramatic income growth: hence the title of the paper. Those who fare less well drop out. Younger performers, in hopes of making it big, perform to small audiences and earn net returns below what they could earn outside the industry—starving artists, etc. The distribution of earnings in the occupation must be positively skewed. On the consumer side, more discriminating members of the audience attend performances by older, established performers, while the less discriminating either see younger, untried performers, or stay home if they are sufficiently indiscriminating.

Comparing steady states yields some unfamiliar results which arise because of the intertemporal linkages on the supply side. To illustrate, an increase in the variable cost of serving audiences may well raise audience size for young performers. The explanation is that one of the benefits gleaned from being a young performer is the value of the chance of becoming a star. Under some circumstances, the cost change can be disastrous for Superstars, who serve a big audience, in which case the chance at stardom is worth less to the young. Thus to attract the young to the industry when costs are higher, they must fare better while they are young, which may mean a much higher price for tickets to their performances and a larger audience for each young performer despite the cost increase.

The balance of the paper sets out and analyzes the model in greater detail. The model is a very simple one and is restricted in many dimensions. Some relaxations are indicated along the way in footnotes. Three more substantive extensions are discussed briefly in Section III.

I. Model

This model is the most austere environment that contains the elements central to the approach being followed.

The agents in the model are performers ($\mathscr{P}$) and consumers ($\mathscr{C}$). Firms, or "promoters," are also present, but most of their behavior is trivial and is therefore suppressed. The behavior of typical $\mathscr{P}$ and $\mathscr{C}$ is first set out and a steady-state competitive equilibrium constructed. In the section following, the main features of the model's equilibrium are discussed along with the impact of altering some of the exogenous parameters.

A. *An Individual Performer*

Performers ($\mathscr{P}$) are risk-neutral maximizers of expected wealth. They may perform in either or both of two periods.[1]

Should $\mathscr{P}$ decline to perform in any period, he participates in some other activity, the value of which is $\bar{w}$ (that may include the value of the difference in utilities associated with performing versus not doing so).[2]

When $\mathscr{P}$ performs, his rendering of the program is either unambiguously good (g) or bad (b). Which of these outcomes occurs is stochastic, and common knowledge. Irrespective of age, performers having no track record are all regarded (by everyone, themselves included) as equally likely to produce a good performance, the probability of this outcome being $p \in (0,1)$.

This structure does not imply that all $\mathscr{P}$ are in fact identical, but rather that the current information on an untried $\mathscr{P}$ is the same for all agents. Thus, for example, if there are n levels of unobservable talent, indexed by i, with the probability of a good performance being θ_i ($i = 1, \ldots, n; \theta_1 < \cdots < \theta_n$) and the fraction of the untried $\mathscr{P}$ who have talent level i being f_i, then

$$p = \sum_{i=1}^{n} f_i \theta_i.$$

Those $\mathscr{P}$ who have performed in the past have a track record that is assumed to be useful for predicting their subsequent behavior. Those who gave a good (bad) performance in the initial period will do so in the current period with probability $p_g > p$ ($p_b < p$). In the example given above,

$$p_g = (1/p) \sum_{i=1}^{n} \theta_i^2 f_i,$$

where $p_g > p$ follows immediately from Jensen's inequality. Serially correlated luck with no underlying heterogeneity also yields the same specification, as does "state dependence," in which giving a good performance actually makes it more likely that $\mathscr{P}$ will do so on the next attempt.

Assume ticket prices to see performances by $\mathscr{P}$ having the various (including null) track records are t_g, t, and t_b (corresponding to p_g, p, and p_b), with $t_g > t > t_b$. This pattern of prices will turn out to be consistent with market equilibrium.

Assuming free entry to the concert promotion business, in equilibrium $\mathscr{P}$ will necessarily receive the excess of gate receipts over the other costs of putting on the show. The most straightforward route is thus to suppose that performers undertake this activity themselves and receive the net revenue from doing so. Thus assume that the cost of organizing the performance, advertising, selling tickets, and hall rental, etc., can be represented as $F + c(a)$, where $F > 0$ is a fixed cost including the cost of advertising, etc., $c(a)$ is the variable-cost function, and a is the size of the audience. For simplicity, let $c(a) = \alpha a^2/2$ for some parameter $\alpha > 0$.

For an untried $\mathscr{P}$ then, net revenue given a is

$$ta - F - \alpha\frac{a^2}{2},$$

in which case the profit-maximal audience size is

$$A \equiv \frac{t}{\alpha},$$

yielding net revenue

$$\pi \equiv t^2/2\alpha - F.$$

Similarly, $\mathscr{P}$ who performed well (poorly) earlier serve an audience of size $A_g \equiv t_g/\alpha$

---

[1] The key features of equilibrium do not depend on the two-period assumption. An m-period version, along the lines studied by myself (1982), is not difficult, although notationally tedious.

[2] $\bar{w}$ does not depend on information obtained by performing. This assumption is not crucial, and may be relaxed so long as alternative opportunities are less responsive to such information in comparison to the value of performing. If this restriction fails, the model may still be analyzed, but the performing activity turns out to be the province of the young or the unsuccessful, something like the minor leagues.

(resp. $A_b \equiv t_b/\alpha$) and obtain net revenue.

$$\pi_g \equiv t_g^2/2\alpha - F \qquad \left(\text{resp. } \pi_b \equiv t_b^2/2\alpha - F\right).$$

As indicated above, $\mathscr{P}$ are assumed to be risk-neutral maximizers of expected wealth. Thus, one who might have performed in the first period, but chose not to, will perform in the second depending on whether

$$\pi \gtreqless \bar{w}.$$

Similarly, for those who did perform, the decision involves one of the comparisons

$$\pi_g \gtreqless \bar{w}, \quad \text{or} \quad \pi_b \gtreqless \bar{w},$$

depending on the first-period reviews. Thus the issue of whether to perform in the first period turns on a comparison of the expected discounted value of performing when young and proceeding optimally thereafter, with the expected discounted value of first choosing the alternative occupation and then proceeding optimally,

$$\begin{aligned} (1) \quad & \pi + \rho\left[p \max\{\pi_g, \bar{w}\}\right. \\ & \left. + (1-p)\max\{\pi_b, \bar{w}\}\right] \\ & \gtreqless \bar{w} + \rho \max\{\pi, \bar{w}\}, \end{aligned}$$

where ρ is the discount factor; $\rho \in (0,1)$.

B. *The Supply of Performances*

It is assumed that there is free entry of untried $\mathscr{P}$ in any period, and that performing is the only way to obtain a track record.[3] A consequence of the first restriction is $\pi \leq \bar{w}$, for otherwise arbitrarily many untried $\mathscr{P}$ would plan to perform in their second period. Moreover, $\pi = w$ is ruled out as well, since $t_g > t > t_g$ implies $\pi_g > \pi > \pi_b$, in which case $\pi = w$ yields (1) as

$$(1') \qquad p\pi_g + (1-p)\bar{w} > \bar{w}.$$

If (1′) held, arbitrarily many young untried $\mathscr{P}$ would plan to enter. Thus

$$(2) \qquad \pi < \bar{w}$$

must be part of any equilibrium. Therefore, any $\mathscr{P}$ who chose not to perform at the first opportunity will never do so. Henceforth, "young" will replace "untried."

Next, given (2), $\pi_b < \pi$ implies $\pi_b < \bar{w}$, so any $\mathscr{P}$ who performed badly in the first period will not perform at all in the second. Thus, the condition implied by free entry is that a young $\mathscr{P}$ must be indifferent between *i*) the present value of permanent attachment to his alternative, and *ii*) the expected present value of performing in the first period and doing so again if and only if he receives good reviews (otherwise taking his alternative):

$$(3) \quad \pi + \rho\left[p\pi_g + (1-p)\bar{w}\right] = \bar{w}(1+\rho).$$

Equation (3) provides further intuition for $\pi < \bar{w}$. Expected (at time of entry) second-period income $p\pi_g + (1-p)\bar{w}$ exceeds $\bar{w}$ because the impact of a bad performance early on can be ameliorated by quitting, thereby obtaining $\bar{w}$. Free entry thus requires a first-period income π smaller than $\bar{w}$, and $\bar{w} - \pi$ is the "option value" of being in the performing industry.

Finally, for any t and t_g satisfying (3) with $\pi < \bar{w}$, young $\mathscr{P}$ will be indifferent about initial entry (and strictly prefer to perform again when older if they earned good reviews earlier). As a consequence, demand determines the number of young $\mathscr{P}$, say N. The total audience served by all N young $\mathscr{P}$ is therefore

$$(4) \qquad NA = \frac{Nt}{\alpha}.$$

---

[3] It is not difficult to permit potential $\mathscr{P}$ to have different alternative opportunities, implying a rising supply price to the industry. $\bar{w}$ is replaced by $\bar{w}(x)$, where $x \in [0, M]$ indexes potential performers within a cohort of size M and $\bar{w}' \geq 0$. Assuming the number of potential performers who enter is $N \in (0, M)$, $\bar{w}(N)$ replaces $\bar{w}$ in (11), below. The only notable departure from what follows is the obvious one that if some $\mathscr{P}$ has poor enough alternatives ($w(0)$ is small), even poor reviews may not cause him to stop performing.

The fraction p of the N young $\mathcal{P}$ receive good reviews. Thus, there are pN older $\mathcal{P}$ serving an aggregate audience of size

$$pNA_g = pNt_g/\alpha. \tag{5}$$

C. *An Individual Consumer*

Every member of the audience must decide whether to see a performance in any period, and what type to attend. This decision is assumed not to depend on either expenditures on other goods or the quality of past performances attended (if there were any).

Members of the audience obtain utility v^0 (measured in dollars) simply by attending a performance, and extra utility v if the performance is a good one. For simplicity, $v^0 = 0$ is assumed for all $\mathcal{C}$. Assuming risk neutrality, $\mathcal{C}$ will choose the type of performance that solves

$$\max\{p_b v - t_b, pv - t, p_g v - t_g, 0\}, \tag{6}$$

where 0 is the value of the null performance "stay home." Recalling that free entry precludes performances by those who received bad reviews earlier on, the element $p_b v - t_b$ may be ignored.

D. *The Demand for Tickets*

Consumers are heterogeneous in terms of the extra utility from a good performance, v. It is assumed that there is a continuum of consumers with v distributed uniformly on $[\underline{V}, \overline{V}]$ and $0 < \underline{V} < \overline{V} < \infty$. Define $\Delta \equiv \overline{V} - \underline{V}$.[4]

In what follows attention will be confined to equilibria in which there is a positive demand for performances by both young and old $\mathcal{P}$ and not all $\mathcal{C}$ buy tickets. Both restrictions merely serve to eliminate repetitive discussion.

The implications of these two assumptions are as follows. First, observe that the net utilities $pv - t$ and $p_g v - t_g$ are both increasing in v. Therefore, if there are to be any $\mathcal{C}$ who do not buy tickets there must be some value of v, denoted, $\underline{v}$, for which the best alternative to staying home is only as good as staying home

$$\max\{p\underline{v} - t, p_g\underline{v} - t_g\} = 0,$$

with $\underline{v} \in (\underline{V}, \overline{V})$. $\mathcal{C}$ for whom $v \leq \underline{v}$ then do not buy tickets, and all other $\mathcal{C}$ attend some performance.

Next, if tickets to see performances by both young and old $\mathcal{P}$ are to be demanded by some $\mathcal{C}$, there must be a value of v, written $\bar{v}$, for which either type of performance is viewed as equally rewarding on net

$$p_g\bar{v} - t_g = p\bar{v} - t,$$

with $\bar{v} \in (\underline{V}, \overline{V})$. Because $(p_g v - t_g) - (pv - t)$ is increasing in v, $\mathcal{C}$ for whom $v \leq \bar{v}$ value performances by the young at least as much as those by the old, and conversely for $v > \bar{v}$. Requiring $\bar{v} > \underline{V}$ is not enough to ensure that there is demand for performances by the young; $\bar{v} > \underline{v}$ is needed as well. If, to the contrary, $\bar{v} \leq \underline{v}$ held, while $\mathcal{C}$ for whom $v < \bar{v}$ would prefer purchasing a ticket to a performance by a young $\mathcal{P}$ to buying a ticket to see an old $\mathcal{P}$, these $\mathcal{C}$ would all rather stay home, and there would be no demand for performances by young $\mathcal{P}$.

Overall, then, taken together, the assumptions that some consumers stay home, and that there is demand for performances by all

---

[4] Much algebra is avoided by requiring v^0 to take on the same value for all $\mathcal{C}$ ($v^0 = 0$) and permitting heterogeneity in v alone. By doing so, the analysis is not altered in any important way unless, in the more general setting, most consumers are relatively indiscriminating (small v). In this case the ability to perform well matters little, and the key condition $t_g/t > p_g/p$, obtained below, can fail. In this case, the dimension along which $\mathcal{P}$ vary is almost economically irrelevant. The situation analyzed in the text, being a simplified version of the instance in which the manner in which $\mathcal{P}$ differ is of some economic importance, is then arguably the one to consider. Also, note that staying home yields the same utility as a bad, but gratis, performance. This restriction may also be relaxed. Finally, allowing v to be distributed nonuniformly is a straightforward extension. The only noteworthy change is that the comparison of steady states is slightly more complicated.

$\mathscr{P}$ imply

$$\underline{V} < \underline{v} < \bar{v} < \bar{V}.$$

Observe that since $\bar{v} = (t_g - t)/(p_g - p)$ and $\underline{v} = t/p$, $\bar{v} > \underline{v}$ gives

$$\frac{t_g}{t} > \frac{p_g}{p}.$$

That is, relative ticket prices t_g/t must exceed relative "talent" p_g/p. Again, should this condition fail, every $\mathscr{C}$ who plans to see a performance would prefer to see a more experienced performer, and there would be no demand for performances by the young.[5]

Given the distribution of v in the population, demand for tickets to see young $\mathscr{P}$ is given by the number of $\mathscr{C}$ for whom $v \in [\underline{v}, \bar{v}]$:

$$\int_{\underline{v}}^{\bar{v}} \frac{1}{\Delta} dv = \frac{1}{\Delta}\left[\frac{t_g - t}{p_g - p} - \frac{t}{p}\right], \tag{7}$$

while demand for tickets to see older, successful performers, is the number of $\mathscr{C}$ for whom $v \in (\bar{v}, \bar{V}]$:

$$\int_{\bar{v}}^{\bar{V}} \frac{1}{\Delta} dv = \frac{1}{\Delta}\left[\bar{V} - \frac{t_g - t}{p_g - p}\right]. \tag{8}$$

E. *Equilibrium*

Collecting the information obtained so far, steady-state equilibrium is characterized by the three equations

$$\frac{1}{\Delta}\left[\bar{V} - \frac{t_g - t}{p_g - p}\right] = \frac{pNt_g}{\alpha}, \tag{9}$$

$$\frac{1}{\Delta}\left[\frac{t_g - t}{p_g - p} - \frac{t}{p}\right] = \frac{Nt}{\alpha}, \tag{10}$$

and

$$\rho p t_g^2 + t^2 = 2\alpha(\bar{w} + F)(1 + \rho p). \tag{11}$$

Equations (9) and (10) equate supply and demand in the markets for tickets to see old and young $\mathscr{P}$, respectively, and (11) restates the free-entry condition (3), with substitution for π and π_g in terms of t and t_g. Variables to be determined are the ticket prices t_g and t, and the number of young $\mathscr{P}$, N.

The system (9)–(11) does not always have a solution with $N > 0$. The problem is as follows. It is trivial to show that for any N, (9) and (10) always have a solution, say $t(N)$ and $t_g(N)$. Moreover both $t'(N) < 0$ and

$$t_g'(N) < 0,$$

with $\lim_{N \to \infty} t(N) = 0 = \lim_{N \to \infty} t_g(N)$,

$$\lim_{N \to 0} t(N) = p\bar{V}, \quad \text{and} \quad \lim_{N \to 0} t_g(N) = p_g\bar{V}.$$

If, for $N \to 0$, t and t_g do not rise enough to satisfy (11), the model has no equilibrium in which any $\mathscr{P}$ ever perform. In this case the alternative ($\bar{w}$) is too good for performing ever to attract any young $\mathscr{P}$. It is therefore assumed that the model's parameters satisfy

$$\rho p (p_g\bar{V})^2 + (p\bar{V})^2 > 2\alpha(\bar{w} + F)(1 + \rho p). \tag{12}$$

A second issue concerns the requirement that not all $\mathscr{C}$ choose to attend some performance. If $\bar{w}$ is sufficiently low, or $\underline{V}$ high enough, all $\mathscr{C}$ will plan to attend some performance given ticket prices consistent with free entry. This outcome is not troublesome except insofar as (10) must be replaced by

$$\frac{1}{\Delta}\left[\frac{t_g - t}{p_g - p} - \underline{V}\right] = \frac{Nt}{\alpha}, \tag{10'}$$

and the analysis carried out in terms of (9), (10′), and (11). In order to remain within the setting characterized by (9), (10), and (11), a slightly more complicated parameter restriction is required. Again this restriction merely states that the alternative $\bar{w}$ is not so poor, or $\underline{V}$ so large, that given ticket prices consistent with free entry all $\mathscr{C}$ would wish to attend a performance.

---

[5] The conclusion $t_g/t > p_g/p$ does not rely on $\underline{v} > \underline{V}$; that is, that some $\mathscr{C}$ do not attend any performance. If all $\mathscr{C}$ did attend, $\bar{v} > \underline{v}$ must be replaced by the condition $\bar{v} > \underline{V}$. However, that all $\mathscr{C}$ attend implies $\underline{V} \geq t/p$, in which case $\bar{v} > \underline{v}$ continues to hold.

II. Predictions

Predictions based on the simple model presented above come in two varieties. The most interesting group comprises descriptive features of steady-state equilibrium: its appearance at a moment in time. The other collection of results involves comparisons of steady states as parameters vary.

A. *Descriptive Features of Equilibrium*

On the audience side, those $\mathscr{C}$ who are relatively indiscriminating—in the sense that they obtain little additional utility from a good performance (low v)—do not participate in the market. The more discerning participate, but confine their attention to lower-priced, less well-established $\mathscr{P}$, and see more poor outcomes. The most discriminating $\mathscr{C}$ pay more, attend performances by artists with a solid track record, and witness few low-quality outcomes.

On the performers' side, young $\mathscr{P}$ earn incomes π below what they could elsewhere, $\bar{w}$. This shortfall is the price of access to the stock of reviewers $\mathscr{C}$ and the possibility of stardom, and is required by free entry. Those $\mathscr{P}$ who receive poor reviews perform no more and sell their skills elsewhere where talent as a performer plays a lesser role. Recipients of good reviews remain in the business, and earn incomes in excess of what they could elsewhere. Also, young $\mathscr{P}$ play to smaller audiences ($A < A_g$) and are less costly to see ($t < t_g$). Moreover, ticket prices and audience sizes differ more than the difference in talent—$A_g/A = t_g/t > p_g/p$—and the difference in incomes is more exaggerated still—$\pi_g/\pi > t_g/t$.[6]

Observe that $\pi_g/\pi > p_g/p$ implies that the implicit price of talent—net returns per unit of talent, π_g/p_g and π/p—are rising in talent: $\pi_g/p_g > \pi/p$. In addition, so are the "marginal returns": $(\pi_g - \pi)/(p_g - p) > \pi/p$. These features identify the model as being of the hedonic variety.

In brief then, the model makes strong claims about the basic dynamic features of occupations in which talent matters but is subject to substantial uncertainty at time of entry. Individuals only join the occupation when young, and seem to pay a high price for doing so. Those who fail subsequently change occupations. The successful remain in their chosen occupation, experience dramatic income growth that appears unjustified by their performance alone, and exert greater control over resources.

Market-level relationships are also implied. In particular, because there are N young $\mathscr{P}$, earning π each, and only pN older $\mathscr{P}$, each obtaining $\pi_g > \pi$, the earnings distribution is always positively skewed, both absolutely and in comparison with the distribution of talent, where the latter is induced by the entry and exit behavior of $\mathscr{P}$. That is, the $(1-p)N$ $\mathscr{P}$ who performed badly when young do not appear in this distribution when old. In addition, the ratio of total earnings of older $\mathscr{P}$ as a group, $pN\pi_g$, to that of younger $\mathscr{P}$, $N\pi$, exceeds p_g, as does relative audience size pNA_g/NA.[7]

Consider, as an example, the performing arts. Therein, a randomly chosen young artist is unlikely to give a virtuoso performance (p is small), but established performers are highly reliable (p_g is large). As a consequence, there will be very few stars, and those who succeed will obtain an enormous payoff and be very costly to see. Young hopefuls will be legion, as will poor performances and brief careers. Moreover, as a group, despite their small number, the stars' aggregate earnings and following will be substantial, perhaps actually dominating those of all other performers taken together.

---

[6] $$\frac{\pi_g}{\pi} = \frac{t_g^2/2\alpha - F}{t^2/2\alpha - F} > \frac{t_g^2/2\alpha}{t^2/2\alpha} > \frac{t_g}{t}, \quad \text{since } \frac{t_g}{t} > 1.$$

[7] $pN\pi_g/N\pi = p\pi_g/\pi > p(p_g/p) = p_g$, and $pNA_g/NA = pt_g/t > p_g$.

Other applications, in business and academe, are immediate.

In contrast, an occupation in which performance may be stochastic but past performance is a poor predictor of current performance—$p \approx p_g$—will exhibit very different features. (The alternative occupation, offering $\overline{w}$, may be interpreted as such an opportunity.) In this case, little will depend on age, or indeed performance, all will fare approximately equally in every period, and the chance of leaving the occupation will be slight. The instruction end of the entertainment business fits the $p \approx p_g$ requirement reasonably well, in addition to being the occupation of many young performers who at first sought performance careers.

B. *Comparison of Steady States*

The implications of altering the model's parameters are easy to obtain. Changes in the structure of consumer demand are examined in some detail. Other experiments can be treated analogously, and are discussed less extensively.

The calculations are less cluttered if some notation is introduced. Thus define $\xi \equiv p_g - p$, $y \equiv \xi \Delta N/\alpha$, $\varphi \equiv \rho p$, and $w \equiv \overline{w} + F$. The system (9)–(11) becomes

$$(13) \qquad t_g(1+py) - t = \xi\overline{V},$$

$$(14) \qquad t_g - t(p_g/p + y) = 0,$$

and

$$(15) \qquad \varphi t_g^2 + t^2 = 2\alpha w(1+\varphi).$$

The endogenous variables are now t_g, t, and y. From (14), the interpretation of $y = t_g/t - p_g/p$ is that y is the "relative ticket price-relative talent" spread.

Totally differentiating (13)–(15), and isolating changes in the endogenous variables on the left-hand side gives

$$\begin{pmatrix} 1+py & -1 & pt_g \\ 1 & -(p_g/p+y) & -t \\ 2\varphi t_g & 2t & 0 \end{pmatrix} \begin{pmatrix} dt_g \\ dt \\ dy \end{pmatrix}.$$

Calling the matrix D,

$$|D| = 2tt_g(p+\varphi) + 2\varphi t_g^2(p_g + py) + 2t^2(1+py) > 0,$$

and

$$D^{-1} = \frac{1}{|D|}$$

$$\times \left(\begin{matrix} 2t^2 & 2ptt_g \\ -2\varphi tt_g & -2\varphi pt_g \\ 2[t + \varphi t_t(p_g/p+y)] & -2[t(1+py) + \varphi t_g] \end{matrix} \right.$$

$$\left. \begin{matrix} t + pt_g(p_g/p+y) \\ t(1+py) + pt_g \\ 1 - (1+py)(p_g/p+y) \end{matrix} \right) = \begin{pmatrix} + & + & + \\ - & - & + \\ + & - & - \end{pmatrix}.$$

Exogenous parameters that may be varied yield changes in the demand for performances (via variation in $\overline{V}$ and/or $\underline{V}$), the value of $\mathscr{P}$'s alternative opportunities ($\overline{w}$), the discount factor (ρ), the technology of performances (F and α), the information content of a performance (p_g), and talent in the population (p). These variations are considered in turn.

The Demand Side. Let $\mu \equiv (\overline{V} + \underline{V})/2$. Then, for given $\Delta = \overline{V} - \underline{V}$, an increment to μ represents a spread-preserving increase in the population average value of v. From above,

$$\begin{pmatrix} dt_g/d\mu \\ dt/d\mu \\ dy/d\mu \end{pmatrix} = D^{-1} \begin{pmatrix} \xi \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} >0 \\ <0 \\ >0 \end{pmatrix}.$$

For given ticket prices, raising μ has the sole effect of increasing demand for tickets to see old $\mathscr{P}$. Equilibrium t_g thus rises, and π_g along with it. Because the change in π_g implies that performing when old becomes a more attractive option, free entry dictates that the return to performing when young, π, must fall, this decline being accomplished by lower t. The increase in the relative ticket price-relative talent spread y is immediate.

Other consequences follow readily. Because $A_g = t_g/\alpha$ and $A = t/\alpha$, larger μ yields bigger audiences for old $\mathscr{P}$ and smaller crowds for the young. This quantity adjustment, acting in conjunction with the change in ticket prices, produces a dramatic change in π_g and π. Also, even though there is free entry to the performance industry, that established $\mathscr{P}$ are in fixed supply given the number of young $\mathscr{P}$ implies that the distribution of earnings becomes even more positively skewed when demand rises.

Finally, recall that the number of $\mathscr{C}$ who do not attend any performance at all is $(\underline{v}-\underline{V})/\Delta$, where $\underline{v}=t/p$, and that the number attending performances by young $\mathscr{P}$ is $(\bar{v}-\underline{v})/\Delta$ with $\bar{v}=(t_g-t)/\xi$. The decline in t thus implies that fewer $\mathscr{C}$ choose not to see any performance. The rise in t_g, and hence t_g-t, then yields the result that more $\mathscr{C}$ view performances by young $\mathscr{P}$. Since each young $\mathscr{P}$ serves a smaller audience, the number of young $\mathscr{P}(N)$ necessarily rises. It then follows that the number of old $\mathscr{P}$ (pN) is greater, and because each serves a larger audience, the number of $\mathscr{C}$ attending performances by older $\mathscr{P}$ rises absolutely and in comparison with the number seeing performances by young $\mathscr{P}$.

Overall then, a spread-preserving increase in the pattern of demand raises the total number of tickets sold and the share of the audience served by established performers, and causes the distribution of rewards to become more skewed. The total number of performers must rise, but the young play to smaller audiences at lower ticket prices and earn smaller rewards. In contrast, older performers play to even larger audiences at higher ticket prices, and earn yet greater incomes.

In terms of interpretation, $d\mu>0$ can be thought of as the outcome of consumers, as a group, deciding that the commodity offered by $\mathscr{P}$ is a useful product, perhaps as a result of trial consumption. The changes in the popular music industry over the past three decades fit this scenario well, as does the recent history of game theory in economics.[8]

Next, consider a mean-preserving spread in the pattern of demand, $d\Delta>0$:

$$\begin{pmatrix} dt_g/d\Delta \\ dt/d\Delta \\ dy/d\Delta \end{pmatrix} = D^{-1}\begin{pmatrix} \xi/2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} >0 \\ <0 \\ >0 \end{pmatrix}$$

The effects of an increment to Δ on t_g, t, y, A_g, A, π_g, and π are qualitatively identical to those induced by an increase in μ. The reasoning is simply that both $d\mu>0$ and $d\Delta>0$ raise the average value of v for $\mathscr{C}$ who would attend some performance at initial ticket prices, which, given free entry, implies higher t_g and lower t. The only notable difference is that the impact on the total numbers of performers (and hence on total performances given by the young, etc.) is now ambiguous.[9]

This experiment may be interpreted in much the same way as the previous one, except that instead of all consumers deciding the commodity is a useful one, a significant proportion instead reach the opposite conclusion, so that the population average does not vary importantly.

Last, a pure increase in demand due to an expansion of the number of $\mathscr{C}$ in the relevant population has the standard effect that would be expected given free entry—N rises by the same proportion, and t_g, t, and y all remain unaffected. This result obtains because demand for tickets to see old $\mathscr{P}$, relative to demand for tickets to see the young, is unchanged. The long-term impact of events akin to the baby boom can be analyzed in this fashion.

Overall then, in terms of application of the results on the demand side, the key question is whether the change to be considered yields overall demand growth and/or changes in demand for tickets to see old $\mathcal{P}$ relative to demand for tickets to see the young.

---

[8] This discussion of the effect of augmenting μ abuses the steady-state assumption, but allows a straightforward interpretation. Moreover, it may be defended with a local stability argument. Similar comments apply to all the parameter changes to follow.

[9] The change in the total number of young $\mathscr{P}$ depends on the fraction of the $\mathscr{C}$ population who would attend a performance by older $\mathscr{P}$ at initial ticket prices: $(\bar{V}-\underline{v})/\Delta$. $d\Delta>0$ $(d\bar{V}=d\Delta/2)$ always reduces demand for performances by young $\mathscr{P}$, but raises demand for performances by older $\mathscr{P}$ when $(\bar{V}-\bar{v})/\Delta$ is less than one-half. Consequently, whether the industry becomes more or less attractive to young $\mathscr{P}$, hence whether N rises or falls, depends on the initial value of $(\bar{V}-\bar{v})/\Delta$.

Alternative Opportunities. Any $\mathcal{P}$'s alternative opportunity is valued at $\bar{w}$. An increment to $\bar{w}$ has effects

$$\begin{pmatrix} dt_g/d\bar{w} \\ dt/d\bar{w} \\ dy/d\bar{w} \end{pmatrix} = D^{-1} \begin{pmatrix} 0 \\ 0 \\ 2\alpha(1+\varphi) \end{pmatrix} = \begin{pmatrix} >0 \\ >0 \\ <0 \end{pmatrix}.$$

An improvement in alternative opportunities raises ticket prices for all performances, but lowers the relative ticket price-relative talent spread y. Intuitively, to attract young $\mathcal{P}$, prospects in the performance industry must improve, and better present and future opportunities both make a contribution through higher ticket prices. The decrease in y reflects the contraction in the industry resulting from $d\bar{w} > 0$ being, effectively, a factor price increase. That is $y \equiv \xi \Delta N/\alpha$, in which case $dy/d\bar{w} \propto dN/d\bar{w} < 0$.

The change in ticket prices raises audience size for all $\mathcal{P}$, with the proportional increase being greater for young $\mathcal{P}$ (because $A_g/A = t_g/t$). Again the induced effect on π_g and π is large due to the combination of ticket price and audience size changes. In terms of market-level aggregates, the main implication is that the increase in t implies fewer $\mathcal{C}$ attend performances (since $v = t/p$).

Discount Factor. Changes resulting from an increase in the discount factor ρ are

$$\begin{pmatrix} dt_g/d\rho \\ dt/d\rho \\ dy/d\rho \end{pmatrix} = D^{-1} \begin{pmatrix} 0 \\ 0 \\ (2\alpha w\rho - \rho t_g^2) \end{pmatrix} = \begin{pmatrix} <0 \\ <0 \\ >0 \end{pmatrix},$$

because $\pi_g > w$ implies $2\alpha w\rho - \rho t_g^2 < 0$. Part of the payoff to entering the performance business is the chance of receiving a good review, and thus earning $\pi_g > \bar{w}$. An increment to ρ raises the current value of this prospect, and so has the same qualitative effect as a reduction in $\bar{w}$.

Production Technology. Production technology gives rise to the cost function $F + \alpha a^2/2$, which has the two parameters F and α.

The consequences of changes in F have been derived already, since $\bar{w} = w + F$ and increments to $\bar{w}$ were analyzed above. To reiterate, a higher fixed cost reduces the size of the industry and raises all ticket prices. Audience sizes for all performances must rise, with audiences for performances by young $\mathcal{P}$ being affected most in percentage terms. In total, fewer $\mathcal{C}$ attend.

Variations induced by changes in the efficiency parameter α are

$$\begin{pmatrix} dt_g/d\alpha \\ dt/d\alpha \\ dy/d\alpha \end{pmatrix} = D^{-1} \begin{pmatrix} 0 \\ 0 \\ 2w(1+\varphi) \end{pmatrix} = \begin{pmatrix} >0 \\ >0 \\ <0 \end{pmatrix}.$$

Raising α increases variable costs, and like an increase in $\bar{w}$ or F (see (15)), augments t_g and t, and lowers y. The market-level implications are also qualitatively the same: fewer $\mathcal{C}$ attend performances.

At the level of individual $\mathcal{P}$, the results are more ambiguous. For example, an increase in α may actually raise audience size for the young. (15) can be written

$$\varphi t_g A_g + tA = 2w(1+\varphi),$$

and since both t and t_g increase with α, it is clear that at most one of A and A_g can do so. If one of A_g or A in fact rises, it must be A, since $A_g/A = t_g/t$, which falls when α increases. Intuitively, the smaller scale at which young $\mathcal{P}$ operate implies that an increment to α raises marginal cost less for them than it does for older $\mathcal{P}$. In conjunction with the fact that ticket prices rise relatively more for the young, it follows that if audience size rises for any $\mathcal{P}$, it must do so for young $\mathcal{P}$.

In regard to compensation, π_g and π, (15) dictates that one must rise and the other fall. Since $\pi = tA/2 - F$, when A rises, or falls proportionately less than t rises, it is π which must increase. However, when both audience sizes fall, that $A_g > A$ implies the absolutely higher ticket prices earned by older $\mathcal{P}$ may yield π_g rising instead. A similar ambiguity enters discussion of the number of young and old $\mathscr{P}$.

Information Content of a Performance. The probability that any $\mathscr{P}$ who has performed well (poorly) previously will do so again in the future is $p_g(p_b)$. As indicated above, that $p_g > p > p_b$ holds may reflect talent, serially correlated luck, or state dependence. In any case, p_g (or ξ, which is equivalent, given p) is one index of the quality of the information contained in a good performance.

One point to note at the outset is that if $p = p_b$, the model's equilibrium is unchanged. While $\mathscr{C}$ would be willing to pay the same amount to see $\mathscr{P}$ with null or poor reviews, those having poor reviews would be older, and attaching no value to the option value aspect of performing, would find their alternative employment strictly more attractive. A corollary is that the value of p_b is irrelevant so long as $p_b \leq p$.

Varying p_g yields

$$\begin{pmatrix} dt_g/dp_g \\ dt/dp_g \\ dy/dp_g \end{pmatrix} = D^{-1}\begin{pmatrix} \bar{V} \\ t/p \\ 0 \end{pmatrix} = \begin{pmatrix} >0 \\ <0 \\ \gtrless 0 \end{pmatrix}.$$

Increasing p_g lowers the level of v, $\bar{v}$, for which indifference between seeing either type of $\mathscr{P}$ at given prices occurs. It thus raises demand for tickets to see older $\mathscr{P}$, and lowers demand for tickets to see the young. The increase in t_g and decline in t are immediate.

The impact on y is, in general, indeterminate. y may be written

$$y = \frac{t_g}{t} - \frac{\xi + p}{p},$$

in which case it is clear that whether y rises or falls depends on how responsive t_g and t are to changes in ξ.

The indeterminacy of the change in y does not disturb results on most of the variables of interest. The change in ticket prices raises audience sizes for older $\mathscr{P}$, and lowers them for the younger. These effects are again magnified in the compensations π_g and π.

The main market-level implication is that the number of $\mathscr{C}$ who do not attend performances, $(\underline{v} - \underline{V})/\Delta$, must fall when p_g rises, simply because t declines. The market split between old and young $\mathscr{P}$ may go either way, as can the number of $\mathscr{P}$.

Talent in the Population. The probability with which any untried $\mathscr{P}$ will put on a good show is p. Increments to p induce changes

$$\begin{pmatrix} dt_g/dp \\ dt/dp \\ dy/dp \end{pmatrix} = D^{-1}\begin{pmatrix} -t_g y - \bar{V} \\ -tp_g/p^2 \\ -\rho t_g^2 + 2\alpha w\rho \end{pmatrix} = \begin{pmatrix} <0 \\ \gtrless 0 \\ \gtrless 0 \end{pmatrix}.$$

At given prices, raising p reduces demand for tickets to see old $\mathscr{P}$, and increases supply by allowing more young $\mathscr{P}$ to perform well. Similarly, demand for tickets to see the young $\mathscr{P}$ is augmented. These effects work to lower t_g and raise t, and in the case of t_g, produce a determinate result. The complication surrounding t arises from the free-entry condition. When p is larger, although t_g (and hence π_g) is lower, it is more likely that any young $\mathscr{P}$ will obtain a good review. If this latter effect is sizable, as it may be for some parameter values, the performing industry can become a more attractive one when p is greater, and t have to fall to satisfy the free-entry condition. Thus while the leading case involves an increase in t, this outcome is not inevitable. When it does occur, the results parallel a decline in p_g.

III. Summary and Extensions

In this paper the distribution of the attribute (talent) that is implicitly priced in a competitive market is derived from an information accumulation process. The point of doing so is to place greater structure on the distribution, thereby generating new predictions on the distribution of rewards and other variables of interest. More generally, the model illustrates the kind of gain associated with imposing structure on the distribution of traits in a hedonic model.

The specific model analyzed is a simple stochastic dynamic version of Rosen's Superstars model. Through performing, useful but imperfect information on the likely outcome of a subsequent performance is obtained. In steady-state equilibrium those older performers who have been recipients of good reviews early on stay in the industry, earn large incomes playing to big crowds, and exert control over great quantities of other inputs. Their audience is not often dissatisfied and pays a high price for this assurance. The less fortunate performers leave the industry. Entry only occurs among the young, who earn low incomes playing to small crowds. Their audience is disappointed comparatively often, but spent relatively little to gain admission. Overall, there are few stars in the industry—the older, well-established performers—but as a group they serve a large fraction of the audience and obtain an even larger share of the returns. The distribution of rewards is always positively skewed.

Owing to its simplicity, the model is not hard to modify. One extension permits young performers to enter the industry, and prior to performing (perhaps as a by-product of costly—say time-consuming—training) obtain imperfect but useful private information on their talent. One sequential equilibrium of this game involves those who receive indications that they are not likely very talented "bombing out" prior to performing, and the luckier proceeding just as above; a familiar phenomenon. The reason is simply that only those who know they are likely to do well are willing to pay the price associated with being a young performer.

A second extension involves choice of repertoire. Suppose that some pieces are easy and others hard, but that the consumption value of a good or bad performance does not depend on difficulty per se. Assume though that a hard piece, while less likely to be played well by any performer, differentiates between talents much better than does an easy one. Under this specification, older $\mathscr{P}$ will never perform anything hard because the only thing that might be gained from doing so is new information, which is not useful to older $\mathscr{P}$. However, provided the hard piece is not too difficult (i.e., does not make the chance of being a star too slim), young $\mathscr{P}$ find it to their advantage to play the hard piece because they have the option of opting out should poor reviews be forthcoming.

Finally, the individual production element (one-man firm) may be replaced by a technology in which each individual's performance is complementary with other factors that may be allocated to his use. In equilibrium, workers who have performed well in the past obtain a greater allocation of complementary factors. This outcome supports a pattern of relative wages that overcompensates for performance differences (analogous to $t_g/t > p_g/p$ above) and produces rising stars within firms.

REFERENCES

Johnson, William R., "A Theory of Job Shopping," *Quarterly Journal of Economics*, May 1978, *92*, 261–78.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, May 1982, *50*, 649–70.

MacDonald, Glenn M., "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review*, December 1982, *72*, 1038–55.

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, December 1981, *71*, 845–58.

______, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, January/February 1974, *82*, 34–55.

______, "The Theory of Equalizing Differences," in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Vol. I, New York: North-Holland, 1987.

Heckman, James and Scheinkman, Jose, "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *Review of Economic Studies*, April 1987, *54*, 243–55.

Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach

By ASSAR LINDBECK AND DENNIS J. SNOWER*

We present a theory of involuntary unemployment which explains why the unemployed workers ("outsiders") are unable or unwilling to find jobs even though they are prepared to work for less than the prevailing wages of incumbent workers ("insiders"). The outsiders do not underbid the insiders since, were they to do so, the insiders would withdraw cooperation from them and make their work unpleasant (that is, "harass" them), thereby reducing the productivity and increasing the reservation wages of the underbidders. The resulting labor turnover costs create economic rent which the insiders tap in wage setting and, as a result, involuntary unemployment may arise.

In order to understand how free-market economies may suffer from protracted spells of involuntary unemployment, it is important to explain why underbidding is not a preponderant feature of labor markets. In this context, we take "underbidding" to mean an agreement between a worker and a firm that a particular job be performed at less than the prevailing wage. (Whether the low-wage offer is made by the worker, the firm, or both, is immaterial in this regard.) If underbidding would occur whenever unemployed workers were willing to work for less than the prevailing wages (normalized for any productivity differences), then involuntary unemployment would either disappear or be accompanied by the empirically unobserved phenomenon of persistent wage deflation.

In the absence of government intervention, underbidding failures can be rationalized by showing (a) why firms have no incentive to agree on low-wage bids in the presence of involuntary unemployment, or (b) why workers lack this incentive. The recent theoretical literature on unemployment has pursued both of these routes. The efficiency wage theories (for example, George Akerlof, 1982; Jeremy Bulow and Lawrence Summers, 1986; Jim Malcomson, 1981; Carl Shapiro and Joseph Stiglitz, 1984; and Andrew Weiss, 1980) have focused on route (a), and much of the labor union literature which has bearing on unemployment and layoffs (for example, Ian McDonald and Robert Solow, 1981; Andrew Oswald, 1982) takes route (b).

This paper attempts to provide a rationale for what many people regard as a well-established social norm, namely, that workers should not "steal" jobs from their fellow workers by agreeing to work for lower wages, and that employers should not permit such "job theft."

Our analysis pursues the two routes above in the following way: Route (a): Firms may refuse to replace incumbent employees with workers who wish to underbid because they realistically expect that, if they would do so, the remaining incumbents would withdraw cooperation from the underbidders in the process of production. Route (b): Unemployed workers may not agree to underbid because they realistically expect that, if they thereby succeeded in replacing some incumbents, their personal relations with the remaining incumbents would be unpleasant

---

*Institute for International Economic Studies, University of Stockholm, S-106 91 Stockholm, Sweden, and Birkbeck College, Department of Economics, 7/15 Gresse Street, London W1P 1PA, England, respectively. We are deeply indebted to Alan Manning, Torsten Persson, Lars Svensson, and two anonymous referees for their valuable comments and suggestions.

(i.e., they would be "harassed" by the incumbents).

In the standard literature on the theory of labor markets, harassment has received little attention, while the theory of teams (for example, Armen Alchian and H. Demsetz, 1972; Jacob Marshak and Roy Radner, 1972), recognizes employees' productivities to be interdependent. However, the crucial, distinctive feature of our analysis is that *cooperation and harassment activities do not occur automatically*; *rather*, *they lie within the control of the employees*, *especially the incumbents*.

In describing the causes and consequences of incumbents' cooperation and harassment activities, we adopt an "insider-outsider" approach to the labor market. The basic idea underlying this general approach is that there are labor turnover costs that generate economic rent which incumbent workers ("insiders") manipulate and exploit in the process of wage determination. In doing so, the insiders primarily take account of their own interests. The unemployed workers are disenfranchized in the process of wage negotiation and, as shown below, involuntary unemployment may occur. It is worth emphasizing that the distinction between "insiders and outsiders" is not merely one between employed and unemployed workers, but rather between groups of workers with different employment opportunities: the insiders face more favorable opportunities than the outsiders due to the insiders' ability and willingness to use turnover costs to drive up their wages.

The insider-outsider approach has been outlined in general terms in Assar Lindbeck and Dennis Snower (1984a, 1985b, 1986a, c). Avner Shaked and John Sutton (1984) provide a bargaining rationale (one of many conceivable ones) for insider market power. The approach has been extended to a two-period framework by Robert Solow (1985) and its implications for time dependence of unemployment (i.e., dependence of current unemployment on past unemployment) were developed by Olivier Blanchard and Lawrence Summers, 1986; Niels Gottfries and Henrik Horn, 1986; and Lindbeck and Snower, 1986b. Robert Gregory (1986) has given some preliminary empirical support (from Australia) for the notion that wages are influenced more by firms' internal conditions than by external conditions in the labor market.

In this paper, the insiders are assumed to create a special, potentially important, variety of labor turnover cost by withdrawing cooperation from, and by harassing the entrants who attempt to underbid. As a result, the insiders are able to raise their wages above the market-clearing level without inducing underbidding.

At these wages, the unemployed workers ("outsiders") would prefer to trade places with the insiders (i.e., they would prefer to be employed for insider wages under insider conditions of work rather than to be unemployed), but they do not have this option. They are victims of *discrimination*, because whenever they gain employment through underbidding, they receive less cooperation and more harassment than the insiders do. In fact, the outsiders may be willing to work for sufficiently less than the insider wages so as to compensate the firms for their more limited cooperation skills, but they may nevertheless be unable to find jobs. Given that the outsiders find themselves with lower productivity and higher disutility of work than the insiders, there may exist no wage that *both* induces firms to hire outsiders *and* induces outsiders to work. This is the sense in which involuntary unemployment can arise in our analysis.

In the literature on unemployment theory, the "insider-outsider" approach may be regarded as an alternative (though complementary rather than mutually exclusive) to the "efficiency wage" approach. Whereas the former explains unemployment through insiders' market power which is used to exploit the rent from labor turnover costs in the process of wage determination, the latter approach explains unemployment through asymmetric information and firms' market power in wage determination (see Lindbeck and Snower, 1986c) for a comparison of the two approaches). In the insider-outsider world, the unemployed workers do not engage in underbidding because the insiders prevent them from doing so; in the efficiency-wage world, underbidding does not occur because it is not in the firms' interests. Among the contributions to the efficiency wage literature, George Akerlof's (1982) "gift exchange model" is closest in spirit to our approach here, in that his analysis describes how employers' wage offers may be used to promote their workers' cooperation and effort.

Section I deals with the microeconomic behavior of workers and firms in our model. Section II describes the equilibrium of a single firm and its employees. In Section III, we incorporate this equilibrium in an aggregate analysis of the labor market and examine how involuntary unemployment can occur in this context. Section IV deals with potential objections to our analysis. Finally, Section V contains concluding remarks.

I. The Behavior of Economic Agents

A. *The Underlying Setup*

Though our explanation of involuntary unemployment rests on two distinct, logically independent arguments (one of which focuses on cooperation, the other on harassment), for brevity, the formal model in Sections II and III deals with these arguments simultaneously.

The cooperation and harassment activities, which the insiders use to protect themselves against underbidding, may be defined as follows. "Cooperation" refers to all those activities in which workers help one another in the process of production and thereby raise their productivity. "Harassment" stands for all those activities whereby workers make one another's jobs more disagreeable (primarily by damaging their personal relations) and thereby raise their disutility of work.

In practice, those workers who have spent a long time at their jobs are often more capable of cooperation and harassment than their newly arrived counterparts. We capture this observation roughly by supposing that when workers first enter their firms, they are unable to cooperate with or harass other workers, but after a fixed period of time—call it the "initiation period"—they all gain identical access to these abilities.

Within this context, we identify three homogeneous groups of workers: (*i*) *insiders*, the "experienced" employees who are able to engage in the full range of cooperation and harassment activities; (*ii*) *entrants*, the "inexperienced" employees who have no access to these activities; and (*iii*) *outsiders*, the unemployed workers.

In this section, we build a simple model that captures the role of insiders' cooperation and harassment activities in the formulation of wage and employment decisions within a firm. Our model is based on the following salient structural assumptions:

1. *Wage Decisions*. Each employee's wage is negotiated for one period at a time, where (for simplicity) the length of the period is assumed equal to the initiation period.[1]

2. *Outsiders Are Perfect Competitors for Jobs*. Thus, when an outsider is hired (and thereby turns into an entrant), the person's entrant wage is equal to his reservation wage (for the duration of the initiation period).

3. *Insiders Have Some Market Power*. Each insider sets his wage "individualistically" (taking the strategies of all other agents as given).[2]

4. *Monitoring*. An insider's wage cannot be made contingent on his cooperation and harassment activities, since the firm is unable to monitor these activities directly. (All that the firm can observe is its output and the number of insiders and entrants it employs.[3])

---

[1] In other words, there are no "long-term" wage contracts (extending over the employees' lifetimes). If such contracts were possible and if employees lacked market power on entering the firms, then involuntary unemployment could not exist.

[2] The assumption of unilateral wage setting by insiders is made only for expositional simplicity, as noted in Section II, Part B.

[3] Our conclusions would not be substantively affected if we would make the more general assumption that the firm (through supervision of its employees) monitors the cooperation and harassment activities imperfectly.

5. *Employment Decisions*. These are made unilaterally by the firms.

6. *Sequence of Decisions*. In the first stage of the decision-making process, the insider wage and the cooperation and harassment levels are set, taking into account how these decisions affect employment. The entrant wage is determined as well. In the second stage, the firms make the employment decisions, taking the insider and entrant wages, as well as the cooperation and harassment levels, as given.

B. *The Firm*

Consider a firm that has two variable factors of production: insiders (L_I) and entrants (L_E). Let a_I represent the level of cooperation among insiders (measured as the actual number of insiders divided into the number that would be required to produce the same output in the absence of cooperation among insiders), and let a_E stand for the level of cooperation between insiders and entrants (measured as the actual number of entrants divided into the number that would be required to produce the same output in the absence of cooperation from insiders). We will call a_I and a_E the "labor endowments" of the insiders and entrants, respectively. The firm is assumed to know the levels of these endowments but it cannot observe the cooperation activities of individual workers. We write the firm's production function as $Q = f(a_I \cdot L_I + a_E \cdot L_E)$, $f' > 0$, $f'' < 0$, where Q is the level of output.

Let W be the insider wage and R_E be the entrant wage (which is equal to the entrants' reservation wage). All insiders are identical and receive the same wage, and similarly for entrants. The firm can observe W and R_E, but it cannot observe the harassment activities which are reflected in the level of R_E.

Within the two-stage, decision-making process specified in Section I (with wages, cooperation, and harassment decisions made in the first stage and employment decisions made in the second), the firm's problem is to maximize its profit with respect to L_I and L_E, taking the insider and entrant wages, the overall cooperation and harassment levels, as well as the production function f as given. To present our analysis in the simplest possible way, we assume that the firm has a one-period time horizon.[4]

Let m be the firm's "incumbent work force," that is, it is the stock of insiders carried forward from the past. Since we assume that cooperation and harassment skills are firm-specific and that entrants acquire them only after they go through the initiation period, it is clear that $L_I \leq m$.

Thus, the firm's profit-maximization problem is

$$\text{Maximize} \quad \pi = f(a_I \cdot L_I + a_E \cdot L_E) - W \cdot L_I - R_E \cdot L_E \tag{1}$$

$$\text{subject to } H_I \leq m; \quad L_I,\ L_E \leq 0.$$

Let $\lambda = a_I \cdot L_I + a_E \cdot L_E$ be the firm's effective work force (i.e., its work force in efficiency units of labor). Then the first-order conditions may be expressed as follows:

$$\frac{\partial \pi}{\partial L_I} = a_I \cdot f'(\lambda) - W \geq 0, \tag{2a}$$

$$\frac{\partial \pi}{\partial L_I} \cdot (m - L_I) = 0;$$

$$\frac{\partial \pi}{\partial L_E} = a_E \cdot f'(\lambda) - R_E \leq 0, \tag{2b}$$

$$\frac{\partial \pi}{\partial L_E} \cdot L_E = 0,$$

where we ignore the nonnegativity constraint on L_I and we assume that the firm is able to hire all the entrants it demands at the wage R_E.[5]

---

[4] With regard to our analytical conclusions, this turns out not to be a restrictive assumption. Naturally, if the firm has a multiperiod time horizon, it faces an inherently intertemporal problem, since the entrants hired in one period become insiders in the next. Lindbeck and Snower (1985a) extend our model to a two-period, overlapping-generations setting.

[5] Since $L_I \leq m$, the marginal incumbent may generate positive profit, as shown by the inequality in (2a).

C. *The Workers*

Insider i has the following decision variables:

1) The level of his harassing activity directed at entrants: h_E^i (implicitly assuming that the insider harasses all entrants in equal measure). We assume that the insider does not harass other insiders.[6]

2) The levels of his cooperative activities directed at other insiders and entrants: a_I^i and a_E^i, respectively, defined analogously to a_I and a_E above. (We assume implicitly that he cooperates with all the other insiders in equal measure, and similarly for all the entrants.)

3) The insider i's wage: W^i.

The insider makes these decisions "individualistically," that is, he takes the optimal cooperation, harassment, wage setting, and employment strategies of all other agents as given. Let us examine the role of each of these decision variables in the context of the insider's decision-making problem.

Since we are primarily concerned with the effect of insiders' cooperation and harassment activities on their wages and employment, it is natural to make the simplifying assumption that the insider i's cooperation and harassment activities do not affect his own utility directly, but only indirectly via the wage he is able to achieve.[7] In other words, the insider is assumed to regard the activities h_E^i, a_I^i, and a_I^i as neither desirable nor undesirable per se and therefore (as shown below) he uses them only to support his wage claims.

We specify insider i's utility function quite simply as $\Omega^i = C^i - l^i$, where C_i is his consumption and l^i is his labor (in units of time). Labor is taken to be a discrete activity, with $l = 1$ for an employed worker and $l = 0$ for an unemployed one. We assume that each worker consumes his entire income in each period. For insider i, this means that $C^i = W^i$.

The insider's reservation wage, R_I^i, is defined as that wage (W^i) which makes him indifferent between employment (yielding utility $W^i - 1$) and unemployment (yielding utility 0):

$$R_I^i = R_I = 1, \qquad \text{for all } i. \tag{3a}$$

Let entrant j's utility be $\Omega^j = C^j - l^j - H_E$, where H_E stands for his disutility from being harassed by the insiders.[8] Naturally, we assume that

$$\left(\partial H_E / \partial h_E^i\right) > 0 \quad \text{for any insider } i. \tag{3b}$$

Thus the entrant's reservation wage is

$$R_E^j = R_E = 1 + H_E, \qquad \text{for all } j. \tag{3c}$$

We assume that each insider's harassing activity (h_E^i) is bounded from above and below, so that each entrant's disutility from being harassed (H_E) is also bounded as

$$0 \leq H_E \leq H, \tag{3d}$$

Since $L_E \geq 0$, the marginal potential entrant may generate negative profit, as shown by the inequality in (2b). Finally, we assume that $A \cdot f'(0) > 1$, where A is the upper bound on a_I, as given in (4c), so that it is always profitable to the firm to employ some insiders: $L_I^* > 0$.

[6] This assumption could be derived from more basic postulates. For example, we could assume that each insider finds it disagreeable to harass the other insiders and then show that, in the Nash equilibrium (described below), no insider is able to achieve a higher wage by harassing other insiders. Thus, each insider chooses not to harass the other insiders.

[7] Allowing the insider's cooperation and harassment activities to affect his utility directly has self-evident implications for our results. In practice, of course, this direct utility effect might be positive or negative. Whereas it is usually safe to assume that harassment activities are disagreeable to the harassers, the same cannot be said of cooperation activities. For example, there are direct utility gains from cooperation when an insider prefers to work in cooperation with entrants or other insiders than to work in isolation. On the other hand, direct utility losses from cooperation are conceivable as well, since an insider who cooperates may expend more "effort" than one who does not.

[8] Recall that the entrants, unlike the insiders, are unable to perform harassment activities. Furthermore, note that H_E is the same for every entrant (since we have assumed that each insider harasses all entrants in equal measure).

where H is a nonnegative constant (described in fn. 17, below).

We now turn to the insider i's cooperation activities and specify how they affect the productivities of the other insiders and the entrants. We wish to ensure, quite plausibly, that an insider is able to raise the marginal products of workers by cooperating with them. For this purpose, we make the following two assumptions:

$$\partial a_I/\partial a_I^i,\ \partial a_E/\partial a_E^i > 0, \tag{4a}$$

and

$$0 < \eta_I,\ \eta_E < 1, \tag{4b}$$

where $\eta_I = -(f''/f')\cdot a_I \cdot L_I$ and $\eta_E = -(f''/f')\cdot a_E \cdot L_E$ are the elasticities of the marginal product of labor with respect to the insider and entrant work forces. Thus, when an insider increases his cooperative activity with other insiders and entrants, the marginal products of the insider and entrant work forces rise:

$$\frac{\partial(a_I\cdot f')}{\partial a_I^i} = \frac{\partial a_I}{\partial a_I^i}\cdot\frac{1}{f'}\cdot(1+\eta) > 0,$$

$$\frac{\partial(a_E\cdot f')}{\partial a_E^i} = \frac{\partial a_E}{\partial a_E^i}\frac{1}{f'}\cdot(1+\eta) > 0.$$

Assumption (4a) means that the jobs within the firm are sufficiently interdependent so that an individual insider's cooperation with other insiders or entrants has a significant, positive effect on their labor endowments. This assumption is unnecessary to our analysis whenever insiders can influence a_I and a_E through coordinated (rather than individualistic) activity. Indeed this suggests that, in large plants with little job interdependence, insiders may have a special incentive to form a union. An insider-outsider explanation for the emergence of unions is to be found here.

We let each insider's cooperating activities (a_I^i and a_E^i) be bounded from above and below, so that the labor endowments of insiders and entrants (a_I and a_E) are bounded as follows:

$$1 \le a_I,\ a_E \le A, \tag{4c}$$

where A is a constant greater than unity.[9]

Finally, we turn to the insider's influence over his wage W^i. To reach our qualitative conclusions, we only need to assume that (a) each insider's wage captures at least some of the economic rent generated through his cooperation and harassment activities, and (b) the greater this rent, the higher his wage. These properties hold in a variety of well-known bargaining games (for example, Shaked and Sutton, 1984) and are in accord with commonsense ideas on wage-setting processes. However, to make our exposition as simple as possible, we consider only the extreme case in which each insider sets his own wage unilaterally and individualistically (as noted in Section I). This means that each insider takes the wages and employment of all other insiders as given. Consequently, if he wishes to retain his job, he must set his wage so that the firm has an incentive to employ him in addition to all the other insiders it is employing. In other words, each insider regards himself as the marginal worker in the firm's employment decisions.

The insider i faces two wage-setting options: (i) he may set his wage at some level V^i, which is sufficiently low to ensure his continued employment, or (ii) he may achieve his reservation wage R_I^i by choosing not to be employed. Clearly, the insider will choose the first option only if the maximum achievable wage V^i, denoted by $V_{\max}^i$, is at least as great as the reservation wage; other-

---

[9] The assumption that A is a constant is merely an expositional simplification. Our analysis could be easily be extended to cover the possibility that A is an increasing function of L_I (i.e., the more insiders there are, the greater the potential for cooperative activity). In that case, we would require that the marginal product of insiders (in the Nash equilibrium described below) diminishes as more insiders are hired.

$$\partial(a\cdot f')/\partial L_I = A\cdot f'' + A'\cdot f' < 0.$$

(This implies that the demand curve for insider labor is downward sloping.)

wise, the second option is preferable:

$$W^i = \max\left[R_I, V^i_{\max}\right]. \tag{5a}$$

$V^i_{\max}$ may be inferred from the firm's employment behavior as described by the first-order condition (2a) and (2b). If entrants are not employed, then only (2a) is relevant for determining the maximum achievable wage: $W \leq a_I \cdot f'(\lambda)$ and thus $V^i_{\max}$ is the maximum of $[a_I \cdot f'(\lambda)]$ with respect to a^i_I and a^i_E. On the other hand, if entrants are employed (in addition to insiders), then both (2a) and (2b) are relevant and the first parts of (2a) and (2b) hold as equalities:

$$\begin{aligned} f'(\lambda) &= f'(a_I \cdot m + a_E \cdot L_E) \\ &= V^i/a_I = R_E/a_E, \end{aligned}$$

and thus $V^i_{\max}$ is the maximum of $[(a_I/a_E) \cdot R_E]$ with respect to a^i_I, a^i_E, and h^i_E.

In short, $V^i_{\max}$ may be expressed as follows:

$$V^i_{\max} = \max_{a^i_I, a^i_E, h^i_E} \left[\min\{a_I \cdot f'(a_I \cdot L_I), (a_I/a_E) \cdot R_E\}\right]. \tag{5b}$$

Substituting (5b) into (5a), and recalling that $R_I = 1$ and $R_E = 1 + H_E$, we obtain insider i's wage:

$$W^i = \max_{a^i_I, a^i_E, h^i_E} \left[1, \min\{a_I \cdot f'(a_I \cdot L_I), (a_I/a_E) \cdot (1 + H_E)\}\right]. \tag{5c}$$

II. The Microeconomic Equilibrium: The Firm and Its Employees

We now show how the equilibrium levels of wages, employment, and cooperation and harassment activities are determined through the interaction of a firm and its employees. Our concept of equilibrium (for the two-stage, decision-making process described above) may be specified as follows:

Definition: *In the Nash equilibrium of the firm and its employees, (a) each insider i maximizes his utility with respect to his decision variables* W^i, a^i_I, a^i_E, *and* h^i_E, *taking the strategies of the firm and the other employees as given, and (b) the firm maximizes its profit with respect to its decision variables* L_I *and* L_E, *taking the strategies of its employees as given.*

Let us now turn to the characteristics of this equilibrium.

A. *Cooperation*

Under equilibrium conditions, each insider cooperates fully with other insiders, but does not cooperate with entrants.

Intuitively, the reason is that (a) by cooperating with the other insiders, the insider raises the marginal product of the firm's incumbent work force and is thereby able to achieve a higher wage, and (b) by refusing to cooperate with entrants, the insider reduces the marginal product of the entrant work force, and consequently reduces the number of entrants hired, thereby raising the marginal product of the incumbent work force and achieving a higher wage.

This can be shown formally by deriving the cooperation activities a^i_I and a^i_E, which permit the insider i to earn his maximum achievable wage $V^i_{\max}$. By the $V^i_{\max}$ equation (5b), the insider's optimal (equilibrium) levels of cooperation are $(a^i_I)^* = \max(a^i_I)$ and $(a^i_E)^* = \min(a^i_E)$, so that

$$a_1^* = A \quad \text{and} \quad a_E^* = 1. \tag{6a}$$

B. *Harassment*

Under equilibrium conditions, each insider harasses maximally all workers who enter the firm.

Intuitively, we see that by doing so, the insider maximizes the entrants' reservation wage and thereby discourages them from entering the firm, so that a minimal number of entrants are hired. Thus, the marginal product of the incumbent work force is maximized, so that the insider achieves the highest possible wage.

Formally, the $V^i_{\max}$ equation (5b) implies that the insider's equilibrium level of harassment (which permits him to earn his maximum achievable wage) is $(h^i_E)^* = \max(h^i_E)$, so that

$$\text{(6b)} \qquad H_E^* = H$$

(recalling that, by (3d), H is the upper bound of H_E).

C. *Wage Determination*

Substituting the optimal cooperation levels (6a) and the optimal harassment level (6b) into the W^i equation (5c), we obtain the following wage equation:

$$\text{(7)} \quad (W^i)^* = W^* = \max[1, \min\{A\cdot f'(A\cdot L_I), A\cdot(1+H)\}] \quad \text{for all } i.$$

This equation has a straightforward interpretation. If the insider's reservation wage ($R_I = 1$) falls short of his maximum wage achievable through employment, ($\min\{A\cdot f'(A\cdot L_I), A\cdot(1+H)\}$), then the insider sets his wage with two independent considerations in mind: W^i must be sufficiently low so that (a) the insider does not become unprofitable to the firm, that is,

$$\text{(8a)} \qquad W^i \leq A\cdot f'(A\cdot L_I),$$

(for otherwise he would be dismissed) and (b) the insider remains at least as profitable as the marginal entrant, that is,

$$\text{(8b)} \qquad W^i \leq A\cdot(1+H),$$

(for otherwise he would be replaced by the entrant). We call (8a) the "absolute profitability constraint" (APC) and (8b) the "relative profitability constraint" (RPC) on the insider wage.

D. *Employment Determination*

Whether the insider wage is given by the reservation wage, the APC or the RPC depends on the size of the firm's incumbent work force m. Recall that the firm faces diminishing returns to labor (i.e., $f'' < 0$) and thus the larger the incumbent work force, the lower the incumbent's marginal product. There are three possible scenarios:

(*i*) *A "large" incumbent work force*: Here the incumbent work force is so large that its marginal product is less than the insiders' reservation wage (R_I). In particular, $m > \overline{m}$, where $\overline{m}$ is the "maximum sustainable incumbent work force" (i.e., the largest possible number of incumbents which the firm may have an incentive to employ) and $\overline{m}$ is given by[10]

$$\text{(9)} \qquad A\cdot f'(A\cdot\overline{m}) = 1.$$

When the incumbent work force is greater than its maximum sustainable level ($m > \overline{m}$), it is clear that the firm finds it worthwhile to reduce employment. What remains to be examined is how large the new work force will be and whether some insiders will be replaced by entrants.

To this end, note that in this scenario *the insider wage will be set equal to the reservation wage*:

$$\text{(10a)} \quad W^* = R_I = 1, \qquad \text{for } m > \overline{m}.$$

The reason is that if W were set beneath this level, then some insiders would have an incentive to quit; whereas, if W were above this level, then some insiders would be dismissed even though they prefer employment to unemployment and these insiders would consequently have an incentive to opt for a lower wage.

---

[10] *Proof*. If the firm is not constrained by $L_I \leq m$ in its maximization problem (1), then (by the first-order condition (2a)) its demand for insiders (L_I^*) rises with a_i and falls with W. By (3a) and (4c), $\max(a_I) = A$ and $\min(W) = R_I = 1$. Substituting these values into the insider demand function (2a), as equality, yields (9).

Given that the insider wage is at its minimum level $W=1$, then (by (2a) and (9)) *the firm employs the maximum sustainable incumbent work force*:

$$L_I^* = \overline{m}, \qquad \text{for } m > \overline{m}. \tag{10b}$$

Since the insiders' marginal product is equal to their reservation wage $(A \cdot f'(A \cdot L_I) = 1)$, the marginal product of an entrant (hired in addition to the insiders) must be less than his reservation wage $(f'(A \cdot L_I^*) < 1 + H)$. Thus, *the firm hires no entrants*:[11,12]

$$L_E^* = 0, \qquad \text{for } \underline{m} > \underline{\overline{m}}. \tag{10c}$$

(*ii*) *An "intermediate" incumbent work force*: Here the incumbent work force is (a) small enough so that its marginal product exceeds the insiders' reservation wage, but (b) large enough so that the marginal product of entrants (hired in addition to incumbents) falls short of the entrants' reservation wage. In particular, $\underline{m} \leq m \leq \overline{m}$, where m is the "minimum sustainable incumbent work force" (i.e., the smallest possible number of incumbents that the firm could employ without having an incentive to hire entrants). $\underline{m}$ is given by[13]

$$f'(A \cdot \underline{m}) = 1 + H. \tag{11}$$

Under these circumstances, *the firm hires no entrants*:

$$L_E^* = 0, \qquad \text{for } \underline{m} \leq \underline{m} \leq \overline{m}, \tag{12a}$$

since (by (11)) the marginal product of an entrant is less than the entrant's reservation wage $(f'(A \cdot m) < 1 + H = R_E^*)$.

Consequently, in setting his wage, each insider is constrained not by the need to remain at least as profitable as the entrants (since entrants are never profitable in this scenario), but only by the need to keep his absolute profitability from falling below zero. In other words, the binding constraint on the insider wage (given by the wage equation (7)) is the APC (Constraint (8a)), whereas the RPC (Constraint (8b)) is redundant.[14] This means that *each insider sets his wage equal to the marginal product of the incumbent work force*:

$$W^* = A \cdot f'(A \cdot m) \qquad \text{for } \underline{m} \leq m \leq \overline{m}. \tag{12b}$$

At this wage, *the firm retains all its incumbents*:[15]

$$L_I^* = m, \qquad \text{for } \underline{m} \leq m \leq \overline{m}. \tag{12c}$$

---

[11] The firm has no incentive to replace some (but not all) of its insiders by entrants, since the profit contribution from replacing one insider by one entrant is

$$\begin{aligned} \Phi &= (\partial\pi/\partial L_E) - (\partial\pi/\partial L_I) \\ &= (W - R_E) - (A-1) \cdot f'(\lambda) < 0, \end{aligned}$$

by (2a) and (2b), and

$$(d\Phi/dL_E)|_{dL_E = -dL_I} = (A-1)^2 \cdot f'' < 0.$$

Clearly, this applies to all three scenarios.

[12] The firm has no incentive to replace all of its insides by entrants for the following reason. In the equilibrium for scenario I, the firms's profit is $\pi^*(I) = f[A \cdot g(1/A)] - g(1/A)$, where $g = (f')^{-1}$. If the firm replaced all its insiders by entrants, its profits would be $\hat{\pi} = f[g(1)] - g(1)$ (since entrants do not harass one another and thus $R_E = 1$). $\pi^*(I)$ (which is described by the area under the insider demand curve and above the $W^* = 1$ line in Figure 2A below) is greater than $\hat{\pi}$ (which is described by the area under the entrant demand curve and above the $W^* = 1$ line in the same figure).

[13] *Proof*. If $m = \underline{m}$, then $L_E^* = 0$, provided that $a_E = 1$ and $R_E = 1 + H$. (The reason is that, by (2b), $(\partial\pi/\partial L_E) = f'(A \cdot \underline{m}) - (1 + H) = 0$.) If $m < \underline{m}$, then $L_E^* > 0$, for any feasible a_E and R_E. (The reason is that $(\partial\pi/\partial L_E) = a_E \cdot f'(A \cdot \underline{m}) - R_E > 0$ for any feasible a_E and R_E.)

[14] Formally, this may be shown as follows. Since $m \geq \underline{m}$ then (by 11) $f'(A \cdot m) \leq 1 + H$. Since $m \leq \overline{m}$, then (by (9)) $A \cdot f'(A \cdot m) \geq 1$. Therefore, $1 \leq A \cdot f'(A \cdot m) \leq A \cdot (1 + H)$. By the wage equation (7), we obtain (12b).

[15] The firm has no incentive to replace all of its insiders by entrants, for the following reason. In the equilibrium for scenario II, the firm's profit is $\pi^*(\text{II}) = f(A \cdot m) - A \cdot f'(A \cdot m) \cdot m$. In the equilibrium for scenario III, the firm's profit is $\pi^*(\text{III}) = f[g(1+H)] - A \cdot (1+H) \cdot m - (1+H) \cdot A \cdot \{(1/A) \cdot g(1+H) - m\}$. Observe that in Figures 2A and 2B below, $\pi^*(\text{II})$ and $\pi^*(\text{III})$, respectively, are described by the areas under the insider demand curve, above the $W^* = 1$ line, and to the left of the incumbent work force m (where $\underline{m} \leq m \leq \overline{m}$ for $\pi^*(\text{II})$ and $m < \underline{m}$ for $\pi^*(\text{III})$). Clearly, $\pi^*(\text{II}) > \pi^*(\text{III})$. The upper bound on H, given in fn. 17, implies that $\pi^*(\text{III}) > \hat{\pi}$ (where $\hat{\pi}$ is defined in fn. 12). Consequently, $\pi^*(\text{II}) > \hat{\pi}$.

(*iii*) *A "small" incumbent work force*: Here the incumbent work force is sufficiently small so that the marginal products of both the incumbents and some entrants (hired in addition to the incumbents) exceed their respective reservation wages. In particular, $m \leq \underline{m}$.

Under this scenario, the insiders cannot completely exclude the outsiders from getting jobs (regardless of their cooperation and harassment activities). Thus, each insider must set his wage with a view to his profitability vis-à-vis the entrants, that is, the binding constraint on the insider wage is the RPC (Constraint (8b)).[16] By the wage equation (7), this means that *the insider wage is a markup (by the factor A) over the entrants' equilibrium reservation wage ($R_E^* = 1 + H$)*:

(13a)
$$W^* = A \cdot (1 + H), \qquad \text{for } m \leq \underline{m}.$$

At this wage, the marginal incumbent is just as profitable as the first entrant (hired in addition to the incumbent work force): $(\partial\pi/\partial L_I^*) = Af'(A \cdot m) - W^* = (\partial\pi/\partial L_I) = f'(A \cdot m) - (1 + H)$. Since the incumbent work force is "small," the first entrant generates positive profit; thus, the marginal incumbent does so, too. Consequently, *the firm retains all its incumbents*:

$$\text{(13b)} \qquad L_I^* = m, \qquad \text{for } m < \underline{m},$$

provided (as shown below) that the firm lacks the incentive to replace all its incumbents by entrants.

Moreover, *entrants are hired until their marginal product is brought into equality with their reservation wage*: $A \cdot L_I^* + L_E^* = A \cdot \underline{m}$ (by (11)) and thus

$$\text{(13c)} \qquad L_E^* = A \cdot (\underline{m} - m).$$

We now inquire under what conditions the firm has no incentive to replace all its incumbents by entrants. If the firm were to pursue this replacement strategy, it would encounter a loss and a gain: the loss would arise because the entrants (unlike the incumbents) could not cooperate with one another, and the gain would emerge because the entrants (in the absence of incumbents) would not be subject to harassment. For the loss to exceed the gain, the harassment level H must fall beneath an upper bound, which is easily derived.[17]

E. *The Microeconomic Equilibrium*

Our results above are summarized in the following proposition:

PROPOSITION 1: *For the Nash equilibrium, the insiders' cooperation and harassment activity levels are $a_I^* = A$, $a_E^* = 1$, and $h_E^* = H$, whenever $L_E^* > 0$. Let the firm's incumbent work force (m) be exogenously given. Then the equilibrium wage and employment levels may be characterized as follows*:

$$\text{(I)} \qquad \text{If } m > \overline{m}, \quad \text{then } W^* = 1,$$
$$L_I^* = \overline{m}, \quad L_E^* = 0.$$

Thus, if the incumbent work force (m) is "large," then the insiders receive their reservation wage ($W^* = 1$) and, in response, the firm employs the maximum sustainable incumbent work force ($\overline{m}$) and does not hire any entrants.

$$\text{(II)} \qquad \text{If } \underline{m} < m < \overline{m}, \text{ then } W^* = A \cdot f'(A \cdot m),$$
$$L_I^* = m, \; L_E^* = 0.$$

Thus, if the incumbent work force is "intermediate," then the insiders are paid the marginal product of the incumbent work force (W^*

---

[16] Formally, it follows from $m < \underline{m}$ that $f'(A \cdot m) > 1 + H$. Thus, $A \cdot f'(A \cdot m) > A \cdot (1 + H) > 1$. By the wage equation (7), we obtain (13a).

[17] Thus, we assume that $0 \leq H \leq H^c$, where H^c is the harassment level at which the profit from retaining all the incumbents and additionally hiring the profit-maximizing number of entrants (π^*(III), defined in fn. 15) is equal to the profit from firing all the incumbents and hiring the profit-maximizing number of entrants instead ($\hat{\pi}$, defined in fn. 12). In particular, H^c is a constant implicitly given by $f[g(1 + H^c)] - A \cdot (1 + H^c) \cdot m - (1 + H^c) \cdot A \cdot \{[g(1 + H^c)/A] - m\} = f[g(1)] - g(1)$.

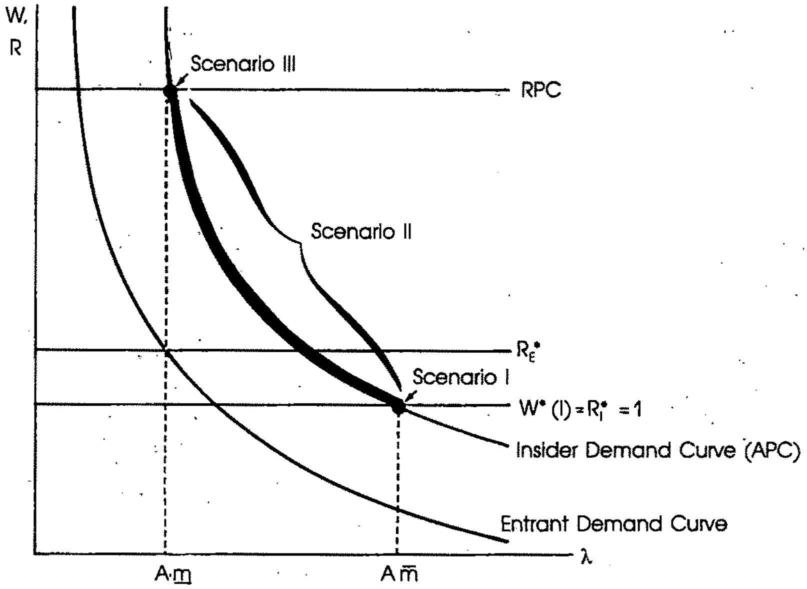


FIGURE 1. THE MICROECONOMIC EQUILIBRIUM; RPC: $W^*(\mathrm{III}) = A\cdot R_E^* = A(1+H)$; INSIDER DEMAND CURVE: $A\cdot f'(\lambda) = W^*$; ENTRANT DEMAND CURVE: $f'(\lambda) = R_E$

$= A - f'(A\cdot m))$ and the firm retains all its incumbents and hires no entrants.

(III) *If* $m < \underline{m}$, then $W^* = A\cdot(1+H)$,

$$R_E^* = 1+H,\ L_I^* = m,\ L_E^* = A\cdot(\underline{m} - m).$$

In other words, if the incumbent work force is "small," then the insider wage is a markup over the reservation wage (with the size of the markup depending on the insider-entrant cooperation differential A) and the firm retains all its incumbents and hires some entrants. The entrants receive the reservation wage, which is raised by the insiders' harassment activity.

Proposition 1 is illustrated in Figure 1. The figure contains two demand curves: an *insider demand curve*, along which the insiders' marginal product is equal to the insider wage, assuming that only insiders are employed: $A\cdot f'(A\cdot L_I) = 0$, and an *entrant demand curve*, along which the entrants' marginal product is equal to the entrant wage (R_E), assuming that only entrants are employed: $f'(L_E) = R_E$. (Thus $f'(A\cdot L_I + L_E) = R_E$ is the demand curve for entrants hired in addition to a given insider work force L_I.)

Observe that the insider demand curve lies above the entrant demand curve (by a factor of A), because the insiders cooperate with one another but are not prepared to cooperate with entrants. The point at which the $R_I^* = 1$ line crosses the insider demand curve yields (by (9)) the maximum sustainable incumbent work force ($A\cdot\bar{m}$, in efficiency units). Similarly, the intersection of the $R_E^* = 1+H$ line and the entrant demand curve yields (by (11)) the minimum sustainable incumbent work force ($A\cdot\underline{m}$, in efficiency units).

The RPC is denoted by the uppermost horizontal line in the figures. The APC coincides with the insider demand curve (since the APC is the locus of wage-employment points at which the absolute profitability of the marginal insider is zero).

In Figure 1, scenario II is depicted by the thick segment along the insider demand curve. In other words, there is a continuum of equilibrium points, each corresponding to a different incumbent work force:

$$\text{(14a)}\quad [W^*(\text{II}), \lambda^*(\text{II})] = [A\cdot f'(A\cdot L_{\text{I}}^*),\ (A\cdot L_I^* + L_E^*)] = [A\cdot f'(A\cdot m), A\cdot m].$$

Here the insiders prevent all entry into the firm through their cooperation and harassment activities and set their wage so as to exploit all their marginal rent ($W^*(\text{II}) = A\cdot f'(A\cdot m)$) and retain their jobs.

Scenario I is pictured by the lowest point on the thick line segment:

$$\text{(14b)}\quad [W^*(I), \lambda^*(I)] = [R_I^*(A\cdot L_I^* + L_E^*)] = [1, A\cdot \bar{m}].$$

Here the incumbent work force is sufficiently large so that the insider wage is reduced to the reservation wage, and, in response, the firm employs only the maximum sustainable incumbent work force.

Finally, scenario III is illustrated by the highest point on the thick line segment:

$$\text{(14c)}\quad [W^*(\text{III}), \lambda^*(\text{III})] = [A\cdot(1+H), (A\cdot L_{\text{I}}^* + L_E^*)] = [A\cdot(1+H), A\cdot \underline{m}].$$

Here the incumbent work force is sufficiently small for some entrants to be profitable at their reservation wage. Thus, the insider wage is set so that the marginal incumbent is just as profitable as the marginal entrant ($W^*(\text{III}) = A\cdot(1+H)$). In response, the firm retains all its incumbents and hires entrants until their marginal product (given by the entrant demand curve in the figure) is equal to their reservation wage (given by the line $R_E^* = 1 + H$ in the figure). Thus the firm's total work force, in efficiency units, is equal to what it would be if the minimum sustainable incumbent work force were employed ($\lambda^*(\text{III}) = A\cdot \underline{m}$).

Figure 1 shows quite simply what the insiders' cooperation and harassment activities are meant to achieve. By cooperating with other insiders, each insider raises the insider demand curve (in the figure) and is thereby able to achieve a higher wage than would otherwise have been possible. This is true for one of two reasons: (*i*) when entrants are not profitable (in scenario II) so that the insider wage is equal to the marginal product of the incumbent work force ($W^*(\text{II}) = A\cdot f'(A\cdot m)$), then cooperation among insiders raises this marginal product; and (*ii*) when entrants are profitable (in scenario III) so that the insider wage is a markup over the entrant wage ($W^*(\text{III}) = A\cdot(1+H)$), then cooperation among insiders raises the firm's cost of replacing an insider by an entrant and thereby increases the markup between the insider wage (W) and the entrant wage (R_E).

When the insider withdraws cooperation from potential entrants, he lowers the potential entrant demand curve (in the figure) and, once again, raises the cost of replacing insiders by entrants.

Finally, when the insider increases his harassment of potential entrants, he raises the entrants' reservation wage, which is the basis on which the insider wage is marked up.

Observe that the insiders' threats to withdraw cooperation and to harass entrants are credible. Given that the firm has already hired a fixed amount, L_E, of entrants, it remains in the insiders' interest to fulfill their threats. For in withdrawing cooperation, each insider causes a reduction in a_E and thereby also reduces the firm's effective work force ($\lambda = a_I^*\cdot L_I^* + a_E\cdot L_E$); as a result, he is able to raise his marginal product and his wage (i.e., for given L_E, $(\partial W_E^i/\partial a_E^i) = a_I\cdot f''\cdot(\partial a_E/\partial a_E^i) < 0$ for any insider i). Furthermore, each insider still has an incentive to fulfill his harassment threat because, in doing so, he raises the entrants' reservation wage and is thereby able to achieve a higher wage for himself.

III. The Aggregate Labor Market: Involuntary Unemployment

We now shift the focus of our attention from the microeconomic equilibrium within a firm to unemployment in the labor market.

Consider an economy that contains a fixed number n of identical firms and a fixed number s of workers. The wage and employment decisions are made in a decentralized fashion within each firm, along the lines indicated in the previous section.

Aggregate labor market activity may be described in terms of three building blocks:

(i) *The aggregate labor demand curve*, denoted by N^D (the thick downward-sloping curve) in Figures 2A–C. (These figures picture the labor market under scenarios I–III, respectively.) When aggregate labor demand (N^D, in efficiency units) is less than or equal to the aggregate incumbent work force ($A\cdot m\cdot n$, also in efficiency units), employment decisions are made along the aggregate insider demand curve:

$$N^D = n\cdot g(W^*/A), \qquad \text{for } 0 \le N^D \le A\cdot m\cdot n, \tag{15a}$$

(by equation (2a), with $a_I = A$), where $g = (f')^{-1}$. Yet when aggregate labor demand exceeds the aggregate incumbent work force, employment decisions are given by the aggregate entrant demand curve:

$$N^D = n\cdot g(R_E^*), \qquad \text{for } N^D > A\cdot m\cdot n \tag{15b}$$

(by equation (2b), when $a_E = 1$).

(ii) The *aggregate labor supply curve*, denoted by N^S (the dashed step function) in Figures 2A–C:

$$W = R_I^*, \qquad \text{for } 0 \le N^{\tilde S} \le A\cdot m, \tag{16a}$$

$$W = R_E^*, \qquad \text{for } A\cdot m < N^{\tilde S} \le \tilde S, \tag{16b}$$

where N^S is measured in terms of efficiency units of labor and $\tilde s$ is the total labor force in efficiency units ($\tilde s = A\cdot m + (s - m)$).

(iii) The *wage-setting curve*, denoted by W^* (the dotted horizontal lines) in Figures 2A–C. When the aggregate incumbent work force is "large" (Figure 2A), insiders receive their reservation wage:

$$W^* = 1, \qquad \text{for } N^D > A\cdot\overline{m}\cdot n. \tag{17a}$$

When this work force is "intermediate" (Figure 2B), workers receive their marginal product:

$$W^* = A\cdot f'(A\cdot m), \qquad \text{for } A\cdot\underline{m}\cdot n \le N^D \le A\cdot\overline{m}\cdot n. \tag{17b}$$

When it is "small" (Figure 2C), workers receive a markup over the entrant wage (which is equal to the entrants' reservation wage):

$$W^* = A\cdot R_E^* = A\cdot(1 + H), \qquad \text{for } 0 \le N^D < A\cdot\underline{m}\cdot n. \tag{17c}$$

Figures 2A–C are drawn so that the labor force $\tilde s$ exceeds the demand for labor at the equilibrium insider wage W^*; thus, some workers remain unemployed. The question to which we now turn is whether this unemployment is *involuntary*.

Clearly, it is not involuntary in scenario I, Figure 2A. Here the wage-setting curve (W^*) passes through the intersection of the aggregate labor demand curve (N^D) and supply curve (N^S). Employment (in efficiency units) curve (N^S) is $A\cdot\overline{m}\cdot n$, leaving ($\tilde s - A\cdot\overline{m}\cdot n$) workers unemployed—voluntarily so, since the incumbents who lose their jobs (($m - \overline{m}$) in number) prefer to be unemployed than to work for their marginal product ($W^* = R_I^* > A\cdot f(A\cdot m)$) and the outsiders (($s - m$) in number) would be unwilling to work at the prevailing wage even if they were not harassed ($R_E^* > R_I^* = W_I{}^*$).

In scenarios II and III, the nature of unemployment is a more complex matter. The reason is that insiders and outsiders in our labor market have different employment opportunities. The difference is twofold:

a) They do not face "identical conditions of employment" (ICE), that is, job attributes lying outside the workers' control. Insiders are able to work under full cooperation and without harassment from the other

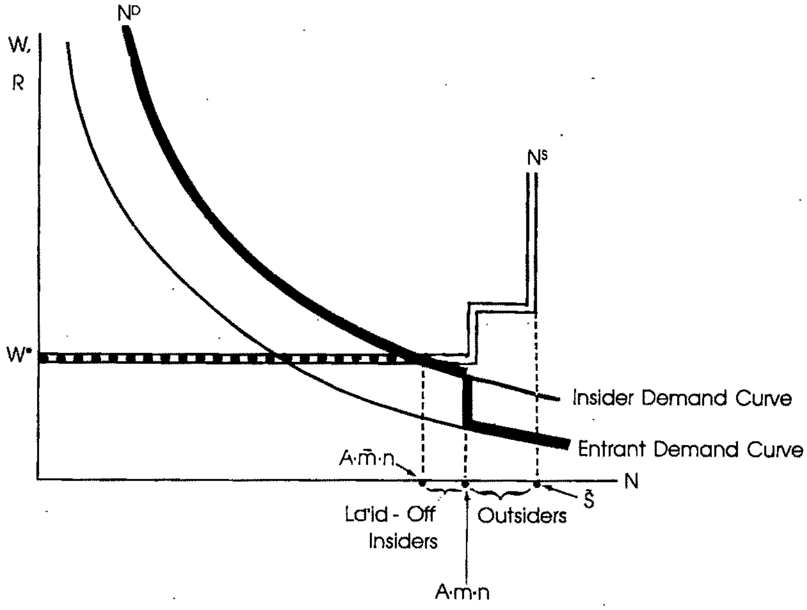


FIGURE 2A. SCENARIO I; N^D: AGGREGATE LABOR DEMAND CURVE (—); N^S: AGGREGATE LABOR SUPPLY CURVE (=); W_* WAGE-SETTING CURVE (●●●●)

insiders, whereas outsiders do not have this option.

b) Even under ICE, insiders and outsiders are not equally productive. Even if insiders and outsiders would receive equal cooperation from their colleagues, the outsiders would still be less productive since they (unlike the insiders) are unable to engage in cooperative activities with others.

These differences suggest that, in order for unemployment to be involuntary, it is not sufficient for workers to be unsuccessful in finding jobs at less than the prevailing wage. Differences in ability should be included in our conception of involuntary unemployment, but differences in conditions of employment (lying beyond the workers' control) should be excluded.

Let the prevailing "efficiency wage" be defined as the prevailing wage normalized for differences in productivity. Then, we propose the following definition of involuntary unemployment:

Definition: *A worker is involuntarily unemployed over a particular period of time if he does not have a job during that period, even though he would wish to work at an efficiency wage which is less than the efficiency wage of a current employee, provided that he had the opportunity to be employed under identical conditions of employment* (ICE) *as that employee.*

This definition is easily extended to a multiperiod context.[18] The definition is meant to

---

[18] For our analysis, the relevant period is the single period over which the firm and the workers optimize their objectives. In a multiperiod context, our definition may be restated in terms of present values: a worker is involuntarily unemployed if he unsuccessfully seeks work (under ICE) for a discounted stream of efficiency wages which is less than the corresponding discounted stream of a current employee, over the same set of time periods. In other words, an outsider is involuntarily unemployed if he has less favorable opportunities than a current employee to earn a present value of wage income for a given stream of productive services. Under our simplifying assumption that outsiders turn into insiders after a single period of work (the initiation period), our intertemporal definition of involuntary un-

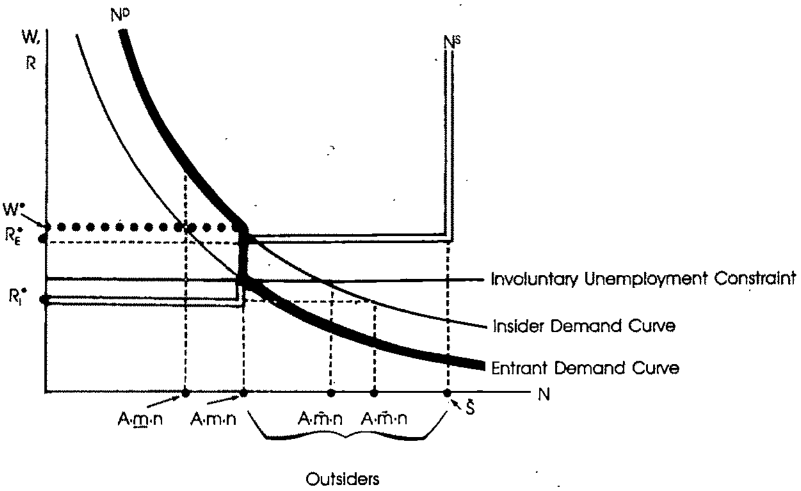


FIGURE 2B. SCENARIO II

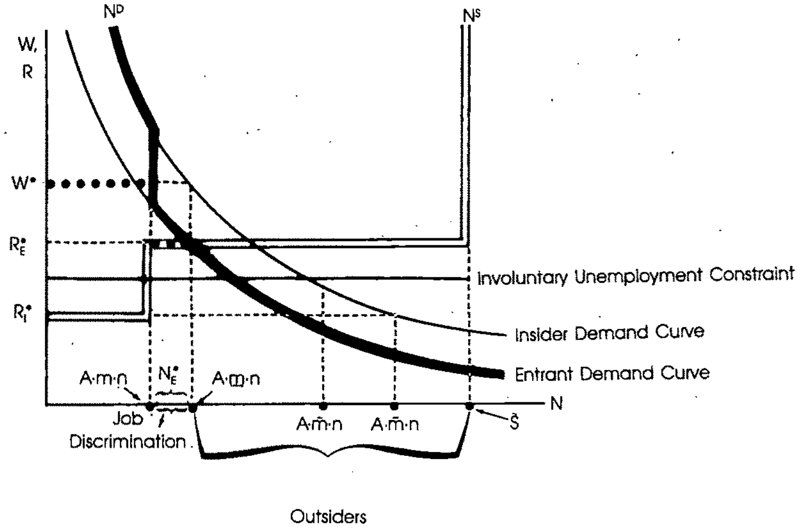


FIGURE 2C. SCENARIO III

capture the idea that outsiders are out of work because they have a smaller choice set —in terms of wages received per efficiency unit of labor—than the insiders. In our model, an insider's efficiency wage is (W/a_I). Under ICE, an outsider's efficiency wage is $(R_E^{\mathrm{ICE}}/a_E^{\mathrm{ICE}})$. If outsiders are involuntarily unemployed, then their choice set is smaller than that of the insiders in the sense that

$$(18a) \qquad \left(R_E^{\mathrm{ICE}}/a_E^{\mathrm{ICE}}\right) < \left(W/a_I\right).$$

---

employment reduces to the condition that the outsider has less favorable opportunities than a current employee *during the initiation period*. Observe that our definition involves a comparison of the wage-labor service opportunities of an outsider and a current employee over a *unique* set of time periods. It is *not* concerned with a comparison of those opportunities over an outsider's future working lifetime with those over an insider's past, present, and future working lifetime. We believe that the notions of unemployment commonly adopted by the news media, politicians, and compilers of unemployment statistics are more readily captured by the former comparison than the latter.

Defining $x \equiv (a_E^{\text{ICE}}/a_I)$ as the ratio of an entrant and insider's labor endowments under identical conditions of work,[19] our condition for involuntary unemployment becomes

$$W > x \cdot R_E^{\text{ICE}} = x \cdot R_I^* = x. \tag{18b}$$

This means that, under identical conditions of employment, an outsider would be cheaper (in terms of efficiency wages) than an insider. (An outsider's labor costs, normalized for productivity differences is $x \cdot R_E^{\text{ICE}}$, whereas an insider's labor cost is W.) The reason why the firms in our analysis nevertheless do not replace insiders by outsiders is that these workers do *not* in fact face identical conditions of employment. Under *actual* conditions of employment (in which insiders receive cooperation and no harassment, whereas potential entrants receive harassment and no cooperation), an outsider is more expensive (in terms of efficiency wages) than an entrant. There is no underbidding because insiders rob their firms of the incentive to employ outsiders.

Our definition of involuntary employment sheds light on the nature of unemployment in scenarios II and III, as shown in Figures 2B and 2C, respectively. Observe that, in both figures, the wage-setting curve (shown by the dotted horizontal line W^*) crosses the N^D curve to the left of the intersection of the N^D and N^s curves. This implies that the outsiders $(\tilde{s} - A \cdot m \cdot n)$ are willing to work for less than the prevailing insider wage (W^*), but are unable to do so. In order for this unemployment to be involuntary, we require that condition (18b) be satisfied, so that the insider wage lies above the "involuntary unemployment constraint," $W^* = x$, in Figures 2B and 2C.[20]

The wage $W = x$ corresponds to a particular size $(\tilde{m})$ of the firm's incumbent work force:

$$x \equiv A \cdot f'(A \cdot \tilde{m}). \tag{19}$$

Thus, we see that the outsiders $(s - A \cdot m \cdot n)$ are involuntarily unemployed only if the aggregate incumbent work force is less than $n \cdot \tilde{m}$. It is easy to show that

$$\underline{m} < \tilde{m} < \overline{m}.^{21} \tag{20}$$

This means that when the aggregate incumbent work force is "small" (scenario I), all the unemployment is involuntary. However, when the aggregate incumbent work force is "intermediate" (scenario II) the unemployment is involuntary when $m < \tilde{m}$ and voluntary when $m \geq \tilde{m}$. (As Figure 2B is drawn, it is involuntary.)

The workers who enter the firm in scenario III face a similar form of discrimination as the outsiders, in that they have a smaller choice set (in wage/efficiency-labor space) than the insiders. In particular, if the entrants and insiders faced identical conditions of employment, then each entrant's compensation per efficiency unit of labor would be less than that of each insider $(R_E^{\text{ICE}}/a_E^{\text{ICE}}) < (W/a_I)$. Hence, in Figure 2C the distance $(A \cdot n \cdot (\overline{m} - m))$ may be called "job discrimination."

Our conclusions concerning the existence of involuntary unemployment are summarized in the following proposition:

PROPOSITION 2: *Consider a labor market described by the aggregate labor demand curve* (15a) *and* (15b), *the aggregate labor supply curve* (16b) *and* (16c), *and the wage-setting curve* (17a) *and* (17b). *Let aggregate incumbent work force*, $m \cdot n$, *be historically given (where n is the number of firms and m is the size of each firm's incumbent work force). If this work force happens to fall short of a*

---

[19] The ratio of the insiders and entrants' marginal products under identical conditions of employment is $[(a_I \cdot f'(\lambda))/(a_E^{\text{ICE}} \cdot f'(\lambda))] = a_I/a_E^{\text{ICE}}$.

[20] Clearly, $1 < x < A$, because the insider is able to engage in cooperative activity whereas the entrant is not. (If insiders and entrants had equal cooperative abilities, then $x = 1$; when insiders cooperate fully with one another but not at all with potential entrants, then $(a_E/a_I) = A$.)

[21] *Proof*. $\tilde{m} < \overline{m}$, since $x > 1$ (by fn. 20) and given the definition of $\tilde{m}$ (in (19)) and $\overline{m}$ (in (9)). Furthermore, $\tilde{m} > \underline{m}$, since $x < A$ (by fn. 20) and given the definitions of $\tilde{m}$ and $\underline{m}$ (in (11)).

particular critical level, $\tilde{m} \cdot n$ *(by* (19)), *then there is involuntary unemployment.*

Although our model deals with the simultaneous performance of cooperation and harassment activities, our explanation of involuntary unemployment may rest on each of these activities alone. If insiders engage in cooperation but no harassment activities, then the equilibrium within the firm is given by Proposition 1 with $H = 0$ and involuntary unemployment arises under the conditions in Proposition 2. When harassment but no cooperation activities are performed, we must not only set $A = 1$ in Proposition 1, but we must also assume that the firm faces some costs of replacing insiders by outsiders. The reason is that, in the absence of such costs, the firm would always find it worthwhile to fire all its harassing insiders and hire entrants who (by assumption) are incapable of harassing.[22]

IV. Some Potential Objections

We now turn to some potential objections to our explanation of wages, employment, and unemployment.

A. *Labor Turnover*

If incumbent workers are able to restrict labor turnover of their firms, why do firms often have large labor turnover rates in practice? Is our analysis inapplicable whenever a firm's work force has large simultaneous inflows and outflows?

When our model is extended to include quits and retirements of employees, it is able to account for simultaneous inflows and outflows. In particular, suppose that $r \cdot m$ randomly chosen[23] incumbents in each firm quit or retire at the end of each time period (where $0 < r < 1$). If the incumbent work force initially exceeds its minimum sustainable level ($m > \underline{m}$), then the separations cause this work force to shrink. As it does so, the insider wage (W) and the marginal product of the insiders and potential entrants rise. However, once the incumbent work force falls beneath $\underline{m}$, the marginal product of potential entrants is so high that the insiders are no longer able to prevent all outsiders from being hired. Consequently, work force inflows and outflows occur simultaneously. In the stationary equilibrium, they are of equal magnitude. It can be shown[24] that these flows are $r \cdot L_I^* = L_E^* = (r \cdot A \cdot n \cdot \underline{m})/(r + A)$, and the associated level of unemployment is $u = s - L_I^* - L_E^* = s - (1 + r) \cdot (A \cdot n \cdot \underline{m})/(r + A)$.

Another reason why insiders' discriminatory activities do not preclude labor turnover in the real world is the following. Suppose that a firm employs "teams" of heterogeneous workers who are complementary to one another in the production process. In this context, insiders have no incentive to prevent the replacement of workers who have quit or retired. On the contrary, since all members of a team are complementary, it is in the insiders' interest to cooperate with and avoid harassing new entrants who fill vacancies on the team. In this light, it becomes clear that our analysis, in the case of heterogeneous labor, applies not to labor turnover within teams, but rather to turnover of teams (namely, the replacement of a team of insiders by a team of entrants).

---

[22] Formally, by the definition of H^c given in fn. 17, if $A = 1$, then $H^c = 0$ and thus, $H = 0$. This implies that the insiders would be unable to erect entry barriers against the outsiders and thus there could be no involuntary unemployment.

[23] The assumption of random choice is made only for expository simplicity, guaranteeing that all workers have the same reservation wage. Had we assumed that workers retire after reaching a particular age, then (a) the reservation wage would rise with age; (b) the firm would hire the youngest entrants available; and (c) it is the reservation wage of these entrants that is relevant to insider wage determination.

[24] Suppose that the incumbent work force is $\hat{m}$ ($< \underline{m}$) when it first falls short of m. In that period of time, the firm hires $L_E^* = A \cdot (m - \hat{m})$. Then, in the next period, the incumbent work force becomes $(1 - r) \cdot \hat{m} + A \cdot (\underline{m} - \hat{m})$. In general, $m_t = (1 - r) \cdot m_{t-1} + A \cdot (\underline{m} - m_{t-1})$. Assuming that $r + A < 1$, the incumbent work force rises monotonically to its stationary level $(A \cdot n \cdot \underline{m})/(r + A)$.

B. *Creation of Firms*

Does our explanation of involuntary unemployment hinge on an assumption that the number of firms in the economy is fixed? Would free entry of firms lead to the elimination of this unemployment?

Our analysis suggests that when insider market power is widespread, entry of new firms (which have no insiders) may be a potentially important route to reducing unemployment. However, it is worth noting that, in practice, the creation of firms is often a lengthy process. Thus, even if free entry would eventually permit full employment to be achieved, the involuntary unemployment may nevertheless last a long time.

Moreover, the existence of involuntary unemployment does not necessarily generate an incentive to create new firms. The mere fact that insiders keep outsiders from being hired by the existing firms does not mean that new firms would find it profitable to hire these outsiders. Observe that new firms, in our analysis, are at a disadvantage vis-à-vis the existing firms, since new firms can employ only entrants (who are unable to cooperate with one another) whereas existing firms also employ insiders (who do cooperate with one another). There are of course many other reasons why new firms may be unprofitable even when existing firms are not, for example, setup costs, capital market imperfections, scarcity of entrepreneurial skills, reduction of product prices due to entry of firms. Hence, the involuntary unemployment of Section IV may persist even after all profitable opportunities for entry of firms have been exhausted.

C. *Output-Related Wage Contracts*

Does our explanation of involuntary unemployment hinge on our assumption that time-rate wages are the only form of labor remuneration? Could output-related wage contracts be used to bribe the insiders not to discriminate against entrants, thereby making the employed and unemployed workers as well as the firms better off and eliminating the unemployment?

Although Pareto-superior alternatives to time-rate wages may exist under some circumstances, they may not be available in others. In fact, they are never available for unemployment generated through differential harassment activities. The reason is that the firm is unable to infer the performance of these activities from the variables it can observe—namely, its total output, its employment of insiders and entrants, and its wages[25]—and thus it has no opportunity to reward insiders for foregoing harassment of entrants.

The matter is not quite so simple for cooperation activities. Although the firm cannot observe these activities directly, it is able to observe its total output. Thus, it may reward its insiders for cooperating with entrants by sharing the proceeds of its output with these insiders. This could take the form of profit or revenue sharing.

Yet, there are a variety of obstacles to designing and implementing such output-related wage contracts. Consider the following three significant ones, which are analyzed formally in Lindbeck and Snower (1985a):

Monitoring-Cost Difficulty. Since profit and revenue-sharing schemes are generally costly for workers to monitor,[26] managers may have an incentive to use their superior position in composing profit or revenue figures to their own advantage. In response, the employees may have an incentive to implement monitoring procedures (and possibly also to engage in litigation). The gains from profit or revenue sharing may not fully compensate the firm and its employees for these monitoring costs.

---

[25] We assume that firms do not know what the outsiders' reservation wages would be in the absence of harassment (for example, they do not know whether these reservation wages are the same as those of the insiders). Thus they cannot infer the presence of harassment by observing the entrant wages (in the event that scenario III obtains).

[26] For instance, managers often have considerable latitude in their choice of profit and revenue accounting practices (for example, how to price intermediate goods and inventories, how to evaluate the firm's debt in real terms, how to treat depreciation and obsolescence).

Risk-Aversion Difficulty. Profit and revenue sharing schemes inevitably involve the imposition of risk on employees. If these employees are risk averse, then they thereby suffer a utility loss. The firm may be unable to compensate them for this loss without robbing itself of the incentive to implement such schemes.

Market-Power Difficulty. When an insider decides to cooperate with entrants, he loses something and gains something: (*i*) he loses market power vis-à-vis the entrants and thus his time-rate wage sinks toward the reservation wage; and (*ii*) he gains some of the profit or revenue which accrues as a result of his cooperation with the entrants. In order for the output-related wage contract to induce insider-outsider cooperation, the second effect must outweight the first. However, that will happen only if the firm relinquishes at least a certain amount of its gross profit. Yet if the firm does so, it may find that its net profit is lower than in the noncooperative equilibrium, and then it has no incentive to implement the contract.

These difficulties, and perhaps others, help explain why output-related wage contracts do not play a particularly prominent role in today's labor markets. However, there is no reason to believe that the difficulties are necessarily insuperable; indeed, the model of differential cooperation activities surely suggests that there is a real-world case to be made for seeking alternatives to time-rate contracts. Be that as it may, time-rate wages are in fact the predominant form of labor remuneration and our analysis indicates how involuntary unemployment may arise when they are used.

D. *Economic Recovery*

Given that insiders can prevent outsiders from getting jobs, does our analysis imply that they can prevent employment from recovering after a recession? In particular, suppose that there has been an upswing in business conditions, shifting the insider and entrant demand curves in Figure 1 to the right. Does our analysis lead to the counterfactual implication that insiders invariably take advantage of such an upswing by raising their wages so that employment remains unchanged?

To see why this potential objection does not hold, let us consider how our macroeconomy responds to an upswing in each of the three scenarios. (Lindbeck and Snower, 1985a, contains a formal analysis of the repercussions of business variations.) To begin with, note that the minimum and maximum sustainable incumbent work force ($\underline{m}$ and $\overline{m}$, respectively) rise in an upswing, so that the dividing lines between the three scenarios in Figure 1 shift to the right. If the incumbent work force is "large" (before and after the upswing), then the insider wage remains at $W^*(I)) = R_I^* = 1$ and more insiders are retained on account of the upswing. If the incumbent work force is "intermediate" (before and after the upswing), the insiders raise their wage ($W^*(\mathrm{II})$) by the full amount of the upward shift of the insider demand curve (without thereby encouraging entry of new employees), and as a result, employment remains unchanged. Finally, if the incumbent work force is "small" (before and after the upswing), the insiders are unable to raise their wage, for otherwise they would induce the firm to replace them by entrants. Consequently, the insider wage remains at $W^*(\mathrm{III}) = A\cdot(1+H)$ and the firm hires more entrants on account of the upswing.

In short, under scenarios I and III, insiders do *not* prevent employment from rising in an upswing, but they do have this effect under Scenario II. In this connection, it is important to mention that if insider wages are assumed to be the outcome of a bargaining process that splits the marginal rent between the insiders and their employers, then an upswing will lead to a rise in employment even under scenario II.

The degree to which an upswing leads to higher wages versus higher employment may depend on the size of this upswing. Consider, for example, a labor market suffering from unemployment and stuck in scenario II. If the upswing is "small," so that the labor market remains in this scenario, then employment will continue to stagnate while insider wages rise. Yet if the upswing is "large," so that the labor market moves into scenario III, then insider wages rise to a particular markup over entrants' reservation wages and employment expands (the larger the upswing, the greater the expansion). Here we observe that a large business stimulus reduces the level of unemployment whereas a small stimulus is unable to do so.

As another example of how the magnitude of the upswing matters to the wage-employment response, consider a labor market in scenario I. Here a "small" upswing (which maintains the existence of scenario I) keeps wages stable and induces firms to fire fewer incumbents than they would have done in the absence of the upswing. Yet if the upswing is large enough to put the labor market into scenario II, then all incumbents are retained and wages rise.

V. Concluding Remarks

This article outlines how insiders' cooperation and harassment activities may give rise to unemployment. The central idea is that firms find it costly to substitute outsiders for insiders, and that insiders manage to capture at least some of the associated economic rent in the process of wage determination. Consequently, insiders raise their wages above the level at which outsiders would be willing to work, but firms nevertheless lack the incentive to replace insiders by outsiders or to add outsiders to their work forces.

In general, the insider-outsider turnover cost can come in many guises, for example, hiring, training, and firing costs (Lindbeck and Snower, 1984a), morale effects of labor turnover (Lindbeck and Snower, 1984b), and this paper explores another one: the insiders' ability and incentive to cooperate with and harass some workers but not others. This ability enables them to create rent and thereby drive up their wages.

In this context involuntary unemployment can arise in the sense that outsiders are unable to find work even though they would be just as profitable to the firm as the insiders, provided that they faced identical conditions of employment. It is the insiders' cooperation and harassment activities which ensure that these conditions are not the same for insiders and outsiders.

Our analysis has a variety of empirical implications. On the whole, these do not square with those of the natural rate hypothesis, since employment within our framework is not uniquely determined by preferences, endowments, and production technologies. Rather, our analysis suggests that the size of the incumbent work force may be an important determinant of employment, since insiders may have market power over wages whereas outsiders do not. As we have seen (in Section II, Part E), past employment (by virtue of its influence on the current incumbent work force) may affect current employment. In particular, over an "intermediate" range of incumbent employment levels ($\underline{m} \leq m \leq \overline{m}$), there is inertia in employment; but this inertia disappears at "low" incumbent employment levels ($m < \underline{m}$) and at "high" ones ($m > \overline{m}$).

Our model also includes the size of the incumbent work force as an argument in the insider wage equation. In particular, over an "intermediate" range of incumbent employment levels, the insider wage is determined in the traditional way, namely, from the relevant marginal productivity condition associated with an estimated labor demand curve (for insiders); yet at "high" and "low" incumbent employment levels, the insider wage is given by the reservation wage (proxied, for example, by unemployment pay, Social Security benefits, etc.) and a markup over the reservation wage, respectively. The reservation wage itself depends on the size of the incumbent work force in our model (since insiders have an incentive to harass other workers only when the incumbent work force is sufficiently small).

Of course, when conducting empirical studies, it is important to take heed of the production technologies under consideration. To the extent that these differ from the diminishing returns to labor assumed in our model, our predictions (regarding the effect of the incumbent work force on employment and wage formation) must be altered accordingly.

Furthermore, our analysis suggests that business upswings will tend to generate fewer jobs in countries with large insider power (due, for example, to unions' ability to exploit cooperation and harassment opportunities) than in countries where insiders are weak. Moreover (as shown in Section IV, Part D), whether a business upswing leads primarily to higher wages or higher employment may depend on the magnitude of this upswing relative to the size of the incumbent work force.

Our model also has implications for cyclical variations in labor market activity. As shown in Lindbeck and Snower (1985a), the movement of wages and employment in an upswing and a downswing may not be symmetric. In particular, a downswing may be characterized by stable insider wages and a contraction of the incumbent work force through retirements and layoffs, while an upswing (as in Section IV, Part D) may take the form of rising insider wages and only modest (if any) increases in employment.

In these and other respects, our insider-outsider approach yields an interrelated set of predictions about labor market activity.

REFERENCES

Akerlof, George A., "Labour Contracts and Partial Gift Exchange," *Quarterly Journal of Economics*, November 1982, *98*, 543–69.

Alchian, Armen A. and Demsetz, H., "Production, Information Costs, and Economic Organization," *American Economic Review*, December 1972, *62*, 777–95.

Blanchard, Olivier and Summers, Lawrence, "Hysteresis and the European Unemployment Problem," *NBER Macroeconomics Annual*, Vol. I, Cambridge: MIT Press, 1986, 15–77.

Bulow, Jeremy I. and Summers, Lawrence H., "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment," *Journal of Labor Economics*, September 1986, *4*, 377–414.

Corden, W. Max, "Taxation, Real Wage Rigidity, and Employment," *Economic Journal*, June 1981, *91*, 309–33.

Gottfries, Niels and Horn, Henrik, "Wage Formation and the Persistency of Unemployment," Seminar Paper No. 347, Institute for International Economic Studies, University of Stockholm, April 1986.

Gregory, Robert, "Wages Policy and Unemployment in Australia," *Economica*, 1986, *53*, 553–74.

Lindbeck, Assar and Snower, Dennis, J., (1984a) "Involuntary Unemployment as an Insider-Outsider Dilemma," Seminar Paper No. 282, Institute for International Economics Studies, University of Stockholm; rev. version, "Wage Rigidity, Union Activity, and Unemployment," in W. Beckerman, ed., *Wage Rigidity and Unemployment*, London: Duckworth, 1986, 97–125.

______ **and** ______, (1984b) "Labor Turnover, Insider Morale, and Involuntary Unemployment," Seminar Paper No. 310, Institute for International Economic Studies, University of Stockholm, 1984.

______ **and** ______, (1985a) "Cooperation, Harassment, and Involuntary Unemployment," Seminar Paper No. 321, Institute for International Economic Studies, University of Stockholm; rev. version as Discussion paper No. DRD257, Washington, DC: World Bank, 1987.

______ **and** ______, (1985b) "Explanations of Unemployment," *Oxford Review of Economic Policy*, Spring 1985, No. 2, *1*, 34–69.

______ **and** ______, (1985a) "Wage Setting, Unemployment, and Insider-Outsider Relations," *American Economic Review Proceedings*, May 1986, *76*, 235–39.

______ **and** ______, (1986b) "Union Activity and Wage-Employment Movements," *European Economic Review Proceedings*, 1986, *31*, 157–67.

______ **and** ______, (1986c) "Efficiency Wages versus Insiders and Outsiders," *European Economic Review Proceedings*, 1986, *31*, 407–16.

Malcomson, Jim M., "Unemployment and the Efficiency Wage Hypothesis," *Economic Journal*, December 1981, *91*, 848–66.

Marshak, Jacob and Radner, Roy, *Economic Theory of Teams*, New Haven: Yale University Press, 1972.

McDonald, Ian M. and Solow, Robert M. "Wage Bargaining and Employment," *American Economic Review*, December 1981, *71*, 896–908.

Oswald, Andrew J., "The Microeconomic Theory of the Trade Union," *Economic Journal*, September 1982, *92*, 576–95.

Shaked, Avner and Sutton, John, "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model," *Econometrica*, November 1984, *52*, 1351–64.

Shapiro, Carl and Stiglitz, Joseph E., "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, June 1984, *74*, 433–44.

Solow, Robert M., "Insiders and Outsiders in Wage Determination," *Scandinavian Journal of Economics*, No. 2, 1985, *87*, 711–28.

Weiss, Andrew, "Job Queues and Layoffs in Labor Markets with Flexible Wages," *Journal of Political Economy*, June 1980, *88*, 526–38.

Bertrand Competition for Inputs and Walrasian Outcomes

By DALE O. STAHL II*

Market making by merchants, who obtain stock from suppliers and resell it to consumers, is modeled as a two-stage pricing game, with winner-take-all competition for the inputs (in contrast to fixed-capacity models). There is a unique subgame-perfect Nash equilibrium (SPNE) which is Walrasian for elastic demand and non-Walrasian for inelastic demand. Alternatively, when merchants first sell forward contracts to consumers and then compete for supplies, the unique SPNE is always Walrasian. Thus, we have an equilibrium model in which the Walrasian price arises not from the benevolent actions of a fictitious auctioneer but from optimal price-setting behavior of merchants.

Consider a model of merchants who obtain stock from suppliers and resell it to consumers, and suppose the merchants face price competition on both sides. Such a model of "Bertrand merchants" is a possible way to endogenize the role of the Walrasian auctioneer in a pure exchange economy. Indeed, a typical story of how markets reach equilibrium involves the notion of arbitraging middlemen. The existing literature on strategic pricing is not really applicable to this model because competition for inputs has not been modeled; either inputs are taken as exogenous or as produced from independent production functions.[1]

This paper fills this void by explicitly modeling competition for the inputs. The notion of "market making" by such Bertrand merchants is modeled as a two-stage game. With one exception, the subgame-perfect Nash equilibrium (SPNE) prices are the Walrasian prices (when SPNE exists). In other words, we have an equilibrium model in which the Walrasian price arises not from the benevolent actions of a fictitious auctioneer but from optimal price-setting behavior of merchants.

These results stand in sharp contrast to the results for capacity-constrained Bertrand competition even with capacity choice. In those models, the outcome is non-Walrasian. The crucial model difference is that in this paper there is "winner-take-all" competition for the inputs, whereas in the other papers either the inputs are taken as fixed or each player has an independent source for capacity production. Thus, the input side of the market is seen as crucial to the overall result.

First, I present a two-stage game in which stocks are acquired in the first stage and sold in the second stage. The supply function and the demand function are nonstochastic and common knowledge to all merchants. Each merchant sets (1) a bid price for the first stage and (2) an ask price for the second stage conditional on everyone's stocks. Sufficient regularity is assumed to ensure a unique Walrasian price, p_c (i.e., the common bid and ask price that equates supply and demand) and a unique "sales-revenue-maximizing price," $\hat{p}$ (i.e., the price a monopolist would charge in the second stage with no capacity constraints). Assume that initial inventories of stock are zero for all merchants. For example, merchants upon waking up in the morning go to the farmers and obtain

---

*Assistant Professor, Department of Economics, Duke University, Durham, NC 27706. This research was conducted in part while the author was on leave at the Department of Economics, MIT, Cambridge, MA 02139. The encouragement and input of Franklin Fisher, Robert Gibbons, Garth Saloner, and Jean Tirole, as well as anonymous referees, is gratefully acknowledged. Naturally the author retains sole responsibility for any errors.

[1] For example, B. Allen and M. Hellwig, 1984, 1986; J-P Benoit and V. Krishna, 1987; W. Brock and J. Scheinkman, 1985; H. Dixon, 1984; D. Kreps and Scheinkman, 1983; E. Maskin, 1986; M. Osborne and C. Pitchik, 1986; and X. Vives, 1986. In another vein, J-P Benassy, 1986; P. Dubey, 1982; and L. Simon, 1984, among others consider strategic behavior on the part of consumers as well as the middlemen.

fresh produce, and then travel to the city to sell this fresh produce to urban consumers. Assuming that the produce perishes overnight, the merchants' inventories would be zero every morning.[2] Also assume that there are no binding capacity constraints on the amount of stock any merchant can acquire during the first stage. The highest bidder acquires all the stock. In the event of equal bids, the supplies are allocated equally among all merchants. In the second stage, consumer sales are rationed to the merchant with the lowest ask price by some mechanism, the result of which determines the residual demand facing the other merchant. This residual demand is assumed to satisfy a mild regularity condition (see Section I).

I will show that there exists an SPNE of this two-stage game with the equal-sharing rule, if and only if the Walrasian price p_c is at least as great as the sales-revenue-maximizing price $\hat{p}$ (equivalently, if and only if the elasticity of demand at the Walrasian price is at least unity). When an SPNE exists, the SPNE bid and ask prices equal p_c, and this Walrasian outcome is the unique SPNE outcome, and all merchants are active in both stages.

Thus, when demand is sufficiently elastic, taking account of both sides of the market resolves the problem with Bertrand price competition first pointed out by F. Edgeworth (1925): namely, that given fixed stocks there need not exist a pure-strategy Nash equilibrium (NE).[3] Only recently has it been shown that even a mixed-strategy NE exists in general for games with Bertrandlike discontinuities (P. Dasgupta and Maskin, 1986), and thus only recently is a resolution technically possible. Despite the fact that the second stage of the merchant game is precisely Bertrand competition with fixed stocks, the Edgeworth problem does not arise when demand is sufficiently elastic because by bidding slightly more, a merchant's stock increases discontinuously, and this has essentially the same effect as if capacity were unlimited.

The nonexistence result when demand is not sufficiently elastic applies to mixed-strategy NE as well as pure-strategy NE. The reason for this nonexistence case is that as long as there are positive monopoly profits for the winner of the first stage, the merchants will outbid one another until they bid up to the "zero-monopoly-rent" bid, $\tilde{p}$, which just reduces monopoly rents to zero. When the demand elasticity is less than unity, the zero-monopoly-rent bid exceeds p_c giving each merchant excess capacity. Then the ensuing competition for customers in the second stage results in revenues less than costs, hence negative profits; but they can guarantee themselves a nonnegative payoff simply by bidding zero. Therefore, the zero-monopoly-rent bid of $\tilde{p}$ is not an NE.

The nonexistence case disappears under a different first-stage allocation scheme for the case of tie bids. Instead of equal sharing of the stocks, if all stocks are allocated to only one of the high bid merchants chosen randomly, then there exists a unique SPNE. In this version of the model, prices are bid up to the zero-monopoly-rent bid $\tilde{p}$, and consumers face the randomly chosen monopolist.

Since this last result applies for all demand elasticities, readers may wonder why I have bothered to develop the equal-sharing version, especially in light of the nonexistence result for inelastic demand. The reason concerns the number of "active" merchants. When both allocation schemes yield unique SPNE prices (namely, p_c), the equal-sharing version has all merchants active in both stages, while the random tie-breaking version has one and only one merchant active in the second stage. Thus, the equal-sharing version has a more realistic outcome. For the cases when equal sharing does not yield an equilibrium, it is observed in Section III that every discrete approximation of the game in which the merchants can choose from only a finite price grid has an SPNE, which as the price grid becomes finer and finer approximates random tie breaking. Thus, nonexistence under the equal-sharing rule can be seen as an artifact of the continuum strategy space. Several other arguments are offered in Section IV that also support using the random tie-breaking version as the proper limit of the equal-sharing version when demand is inelastic while maintaining the equal-sharing rule and its compatibility with many active merchants when demand is elastic.

---

[2] This dynamic story is told solely for the purpose of motivating the zero inventory assumption of the two-stage game. The repeated version of the game will be briefly addressed at the end of the paper.

[3] Suppose two firms each have stock x, which is less than the demand at marginal cost but more than the Cournot quantity. Let p^* be the price at which demand equals $2x$, and suppose p is a pure-strategy NE. For $p > p^*$, undercutting is advantageous. For $p \leq p^*$, raising price will result in positive residual demand and locally increasing revenue. Therefore, p could not have been an NE.

The model considered so far is appropriate for markets in which supplies are obtained first and then sold to consumers, for example, retail food and merchandise. However, in some markets merchants sell forward contracts to customers for delivery, and then go to the suppliers to obtain the needed stocks, for example, farm products sold to food-processing firms, imported goods sold to retail distributors, and a significant portion of wholesale business.

To model merchants in the latter kind of market a second model is considered which is like the first except that the stages are reversed. A penalty for default is imposed sufficient to prevent defaults. *The unique SPNE outcome is Walrasian*; that is, the bid and ask prices equal p_c, and all merchants are active in both stages. The elasticity of demand no longer plays a crucial role. Intuitively, forward contracts remove the "capacity constraint" so the classic Bertrand result holds. It also holds for any number of merchants greater than one.

With or without forward contracting, when the elasticity of demand is at least unity, the outcome is Walrasian with all merchants active. However, for lower elasticities the outcome is sensitive to the institutional aspects of the model. With forward contracting the outcome is still Walrasian with all merchants active. Without forward contracting, the SPNE (of the random tie-breaking version) has non-Walrasian prices and a second-stage monopolist.

These results are dramatically different from the results of the fixed-capacity models and the capacity choice models. In those models the outcome is either monopolistic or Cournot, not Walrasian. The *important modeling difference* is that the latter models involve *no competition for the inputs*, while the Bertrand merchant model has "*winner-take-all*" *competition for the inputs*. The lesson is that the input side of the market can have significant effects on the nature of the output equilibrium. Assuming unrestricted availability of inputs at a constant marginal cost may generate singular results.

The paper is organized as follows. Section I sets out the notation and basic assumptions. Section II analyzes the case of a single monopolist in order to provide a benchmark for comparison. Section III develops the model without forward contracts, and Section IV develops the model with forward contracts. The results are discussed in Section V. Several involved proofs are relegated to an Appendix.

I. Notation and Basic Assumptions

Let (p_{bi}, p_{ai}) denote the bid and ask prices, respectively, of the ith merchant. These prices are the strategic variables of the merchants. Given a bid price p_{bi}, the merchant acquires whatever stock the suppliers are willing to provide at that price. Let x_i denote the stock on hand of the ith merchant after the first stage. Given p_{ai}, the merchant sells whatever stock the consumers are willing to buy (up to x_i).

Let $S(p)$ denote the supply function of the suppliers and let $D(p)$ denote the demand function of the consumers. Assume that both are continuously differentiable with $S' > 0 > D'$. The requirement that the supply function be upward sloping and not vertical will be relaxed in Sections III and IV after the main results are presented.

Assume that there is a unique Walrasian equilibrium price p_c such that $D(p_c) = S(p_c)$. Further assume that the sales revenue function $pD(p)$ is strictly concave, so there is a unique sales-revenue-maximizing price $\hat{p}$.

The ask price subgame (of Section III) with sales constrained by stocks x_i must specify the realized sales given any pair of ask prices (p_{a1}, p_{a2}). If the ask prices are identical, we will assume that sales are split equally among the merchants. If the ask prices differ, say $p_{a1} > p_{a2}$, the low-priced merchant 2 sells his entire stock x_2 or $D(p_{a2})$, whichever is smaller. In the event that $D(p_{a2}) > x_2$, only some of the consumers can be satisfied by merchant 2. Which consumers are satisfied and to what extent is determined by a rationing mechanism. This in turn determines the residual demand facing the high-priced merchant 1. Given merchant ask prices $p_{a1} > p_{a2}$ and equal stocks $x_i = x$, let $\mathrm{RD}(p_{a1}, p_{a2}; x)$ denote the *residual demand* facing the high-priced merchant 1.

The exact form of RD( ) depends on the rationing scheme. Under "uniform rationing" (R. Levitan and M. Shubik, 1972; and Kreps and Scheinkman, 1983), $\mathrm{RD}(p_{a1}, p_{a2}; x) = \max\{D(p_{a1}) - x, 0\}$; that is, a parallel leftward shift of the demand curve. Under "random rationing" (M. J. Beckmann, 1965; and Allen and Hellwig, 1984, 1986),

$$RD(p_{a1}, p_{a2}; x) = (1-\alpha)D(p_{a1}),$$

where $\alpha = \min\{x/D(p_{a2}), 1\}$; that is, a clockwise-inward rotation of the demand curve. Rather than specify one of these or some other arbitrary rationing mechanism, I take the approach of directly assuming that the residual demand function satisfies certain conditions.

Definition: The *residual demand condition* (RDC) *holds if* (i) $\mathrm{RD}(p_{a1}, p_{a2}; x)$ is jointly continuous in (p_{a1}, p_{a2}) and decreasing in p_{a1};

$$\text{(ii)} \quad \lim p_{a1} \to p_{a2} RD(p_{a1}, p_{a2}; x) = \max\{D(p_2) - x, 0\};$$

and (iii) $\mathrm{RD}(p_{a1}, p_{a2}; x) \le D(p_{a1})/2$ for all $p_{a1} > p_{a2} \ge D^{-1}(2x)$.

The continuity condition (i) is natural. Condition (ii) means that for p_{a1} arbitrary close to but above p_{a2}, the residual demand facing the high-priced merchant is approximately the total demand $D(p_{a2})$ less that part of total demand satisfied by the low-priced merchant, $\max\{x, D(p_{a2})\}$. This is a very natural assumption. The substantive part of RDC is condition (iii) which implies [provided $p_{a2} \ge D^{-1}(2x)$] that the residual demand of the high-priced merchant is never more than half the demand he would face alone.

Both uniform and random rationing (and any convex combination) satisfy these conditions.[4] C. Davidson and R. Deneckere (1986) have shown that the uniform rationing scheme produces singular results in the Kreps-Scheinkman model, so it is especially noteworthy that the Bertrand merchant results are robust to the much wider class of rationing mechanisms satisfying the RDC.

II. Benchmark Monopoly Outcome

As a point of reference it is useful to recall how a monopolistic merchant would set prices. This optimization problem will be familiar to every reader if we first convert to quantities as the choice variables. If x is the quantity sold, then clearly the monopolist must acquire at least x and will not want to pay for more. The optimal solution is that x which equates marginal revenue (MR) with marginal cost (MC),

$$\text{(3)} \quad \mathrm{MR}(x) = \partial\left[xD^{-1}(x)\right]/\partial x = \partial\left[xS^{-1}(x)\right]/\partial x = \mathrm{MC}(x),$$

where D^{-1} and S^{-1} are the inverse demand and supply functions, respectively. Letting x_m denote the optimal quantity, the price

---

[4] To see that uniform rationing satisfies RDC, note that if $D^{-1}(2x) \le p_{a2} < p_{a1}$, then $D(p_{a1}) < D(p_{a2}) \le 2x$, so $\mathrm{RD}(p_{a1}, p_{a2}; x) < D(p_{a1})/2 < x$. For random rationing, note that if $D^{-1}(2x) \le p_{a2} < p_{a1}$, then $D(p_{a1}) < D(p_{a2}) \le 2x$, so $\alpha \ge 1/2$ and hence $\mathrm{RD}(p_{a1}, p_{a2}; x) \le D(p_{a1})/2$. The reader may check that RDC (i) and (ii) are also satisfied.

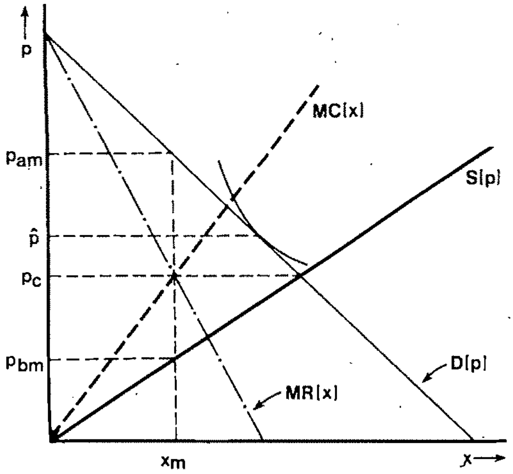


FIGURE 1

solutions are then $p_{am} \equiv D^{-1}(x_m)$ and $p_{bm} \equiv S^{-1}(x_m)$. This is illustrated in Figure 1.

It is obvious that p_{am} is higher and p_{bm} is lower than the Walrasian price p_c. It is also noteworthy that p_{am} is higher and p_{bm} is lower than the sales revenue maximizing price $\hat{p}$.

III. Duopolist Merchants Without Forward Contracts

Let us first consider the case of two merchants and later generalize to $N \geq 2$ merchants. The solution is derived by backward induction. Due to the winner-take-all assumption in the first stage, it will suffice to consider only the case of equal stocks. Given stocks $x_1 = x_2 = x$, the merchants face Bertrand price competition for consumers in the second stage. For any ask prices (p_{a1}, p_{a2}), let $z_i(p_{a1}, p_{a2}; x)$ be the resulting sales for merchant i.

$$z_1(p_{a1}, p_{a2}; x)$$
$$= \begin{cases} \min\{x, D(p_{a1})\}, & \text{if } p_{a1} < p_{a2} \\ \min\{x, D(p_{a1})/2\}, & \text{if } p_{a1} = p_{a2} \\ \min\{x, RD(p_{a1}, p_{a2}; x)\}, & \text{if } p_{a1} > p_{a2}. \end{cases}$$

Similarly for the second merchant. We assume that the residual demand facing the high-priced merchant, $RD(p_{a1}, p_{a2}; x)$, satisfies the residual demand condition (RDC) of Section I.

Let $ER_i(x)$ be the expected revenue from a second stage NE when both merchants have equal stocks $x_i = x$. The following lemma gives the key characterization of these equilibrium expected revenues. Let $M(x) = \max\{pD(p) | D(p) \leq x\}$, the maximum second-stage monopoly revenues attainable with stock x. Note that $M(x)$ implicitly permits a monopolist to dispose of excess stock, which he will do whenever $x > D(\hat{p})$.

LEMMA 1: *Given the RDC,* $ER_i(x) \leq M(2x)/2$ *for all i. (Proof in Appendix.)*

In other words, given the assumption on residual demand and given equal stocks, for each merchant, the expected revenue from a second-stage NE never exceeds the amount they could get by colluding and sharing the monopoly profits equally.

The first-stage acquisitions, given pure strategies (p_{b1}, p_{b2}) and Bertrand competition are denoted by $x_i(p_{b1}, p_{b2})$.

$$x_1(p_{b1}, p_{b2}) = \begin{cases} S(p_{b1}), & \text{if } p_{b1} > p_{b2} \\ S(p_{b1})/2, & \text{if } p_{b1} = p_{b2} \\ 0, & \text{if } p_{b1} < p_{b2}. \end{cases}$$

The outcome is "winner take all," except that in the event of equal bids the merchants share the supplies equally.

It is notationally convenient to define $\mu(p) = M[S(p)] - pS(p)$, the net profits to the winner bidding p (assuming no ties). Then define "zero-monopoly-rent" bid, $\tilde{p}$, as the highest bid price that will yield nonnegative profits for the winner: $\mu(\tilde{p}) = 0$. Clearly no merchant will want to bid higher than $\tilde{p}$.

Figures 2A and B illustrate the two possible cases: the first where $\tilde{p} > p_c$, and the second where $\tilde{p} = p_c$. In Figure 2, the function $g(p_b)$ gives the ask price p_a such that $D(p_a) = S(p_b)$; it is downward sloping because $S' > 0 > D'$. The points of this locus can be read directly from Figure 1. For any quantity, x, the intersections of a vertical

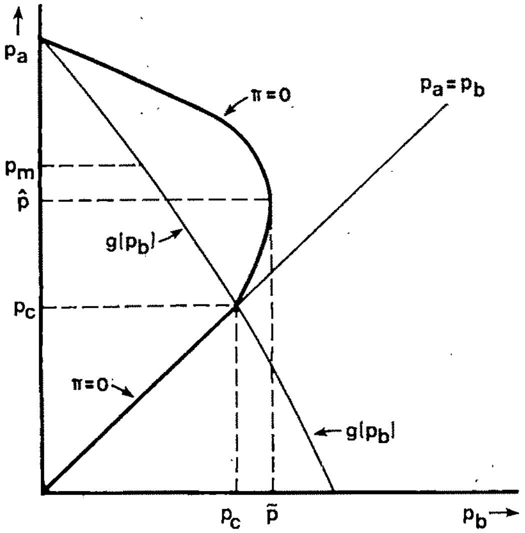


FIGURE 2A

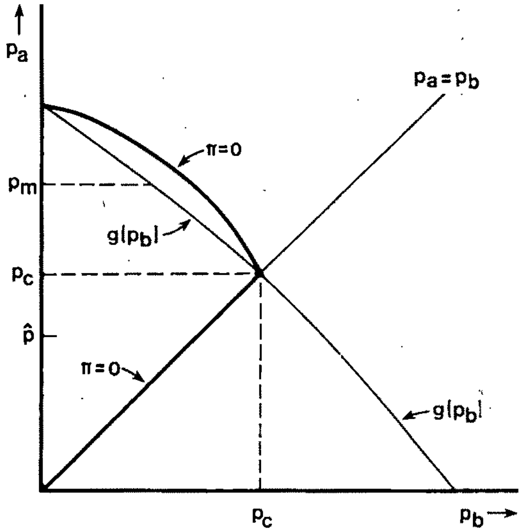


FIGURE 2B

line at x with the supply curve and with the demand curve define a pair of prices (p_b, p_a), which identify a point on the $g(\ )$ locus of Figure 2. Note that everywhere to the right of $g(\ )$, there is excess supply. The zero-profit locus for the winner is the emboldened line lying north and east of $g(\ )$, the portion of the 45° line below p_c, and the vertical axis. In the region surrounded by this locus, the winner makes strictly positive profits; outside this region, he makes strictly negative profits. To see this, note that on the $g(\ )$ locus supply equals demand, so profits are just the common quantity times the ask/bid spread $(p_a - p_b)$, which is positive above the 45° line. Hence, the north and east boundary of the zero profit locus must lie to the north and east of $g(\ )$, and obviously $p_a \geq p_b$.

LEMMA 2: *A necessary condition for an SPNE strategy is that for each merchant the (potentially mixed) strategy be concentrated at the zero-monopoly-rent bid of* $\tilde{p}$.

In terms of pure strategies, given Lemma 1, merchants clearly have the incentive to outbid each other until $p_{bi} = \tilde{p}$. The detailed proof that this holds for all potential mixed strategies is given in the Appendix.

The expected profits when both merchants choose the zero-monopoly-rent bid of $\tilde{p}$ are $E\pi = ER[S(\tilde{p})/2] - \tilde{p}S(\tilde{p})/2$, dropping subscripts because of symmetry. From Lemma 1, $ER(x) \leq M(2x)/2$; hence, we have $E\pi \leq \{M[S(\tilde{p})] - \tilde{p}S(\tilde{p})\}/2 = \mu(\tilde{p})/2 = 0$; that is, expected SPNE profits cannot be positive. Clearly, if these expected profits are negative $(E\pi < 0)$, then no SPNE exists because the merchants can simply bid zero and secure a return of 0.

PROPOSITION 1: *Under the equal-sharing rule for tie bids, if the sales-revenue-maximizing price* $\hat{p}$ *exceeds the Walrasian price* p_c, *then no SPNE exists.*[5]

---

[5] While I believe that SPNE are the most relevant and interesting, it is noteworthy that Lemma 2 and Proposition 1 apply to all NE. To see this recall that the difference between an SPNE and a non-SPNE is that the latter involves strategies off the equilibrium path that are not NE of the subgames and which act as threats affecting the equilibrium outcome. Intuitively, however, there are no effective threats in this merchant game. The only way one merchant can affect the other in the second stage is to have some stock, which entails a bidding tie. But any such threat is nullified by a higher bid by the threatened party unless both are already bidding $\tilde{p}$. But the latter case is moot because neither would want to bid $\tilde{p}$ in the first place. A formal proof could be provided.

PROOF:

As illustrated in Figure 2A, when $\hat{p} > p_c$, then $\tilde{p} > p_c$. Given that both merchants bid $\tilde{p}$ and split the supplies equally, the only way to make nonnegative profits is for both to ask the sales-revenue-maximizing price $\hat{p}$ in the second stage. However, $p_{a1} = p_{a2} = \hat{p}$ is not an NE of the second stage, since with excess capacity each merchant could discretely increase his sales revenues by slightly undercutting the other.

PROPOSITION 2: *Under the equal-sharing rule for tie bids, if the sales-revenue-maximizing price* $\hat{p}$ *does not exceed the Walrasian price* p_c, *then there exists a unique SPNE. The corresponding strategies are* (1) *to bid* $p_{bi} = p_c$, *and* (2) *given first-stage stock acquisitions, to play second-stage NE strategies. In equilibrium, for both merchants,* $p_{ai} = p_{bi} = p_c$ (*which also happens to equal the zero-monopoly-rent bid of* $\tilde{p}$.

PROOF:

As illustrated in Figure 2B, $\hat{p} \le p_c$ implies $\tilde{p} = p_c$. From Lemma 2, the only possible SPNE bids are $p_{bi} = \tilde{p} = p_c$. The unique NE of the second stage is $p_{ai} = p_c$ yielding $E\pi = 0$. To see this, note that in the second stage, $x = S(p_c)/2$, which is strictly less than $D(p_a)/2$ for all $p_a < p_c$. In other words, a merchant would gain no sales by undercutting p_c. Further, since $\hat{p} \le p_c$, residual revenues (given RDC) are decreasing for all $p_a > p_c$. Therefore, there is no incentive to raise prices.

Under the assumption that $pD(p)$ is strictly concave, we can restate Propositions 1 and 2 in terms of the elasticity of demand evaluated at the Walrasian price p_c. Let η_c denote this elasticity (expressed as a nonnegative number). Then, $\hat{p} \le p_c$ iff $\eta_c \ge 1$. In other words, an SPNE exists only for markets with sufficiently high elasticity of demand (namely $\eta_c \ge 1$), in which case the unique SPNE is Walrasian.

Propositions 1 and 2 hold for any number of Bertrand merchants greater than one. (The proofs mimic those given for a duopoly; details are in the Appendix.) Thus, when an SPNE exists, all merchants are active in both stages. That is, they all bid p_c, obtain an equal share of the available supplies, and then sell all their stock at an ask price of p_c.

It should be emphasized that Proposition 1 asserts not only the nonexistence of a pure-strategy SPNE when $\hat{p} > p_c$, but also the nonexistence of even a mixed-strategy SPNE.[6] Recall that any game with a finite strategy space always has a mixed-strategy NE, and an approximation of the model by confining prices to a finite price grid (say units of cents or mils). As the price grid becomes finer, the NE-mixed strategies become more concentrated near the zero-monopoly-rent bid $\tilde{p}$, but in such a manner that the probability of ties tends to vanish.[7] Hence, the outcome tends to look as if one merchant were randomly chosen from among the high bidders, but in the limit game both merchants bid $\tilde{p}$ and share the supplies. Thus, the nonexistence result can be viewed as an artifact of the continuum strategy space, and in low elasticity cases it might be better to interpret the equal-sharing version "as if" there were a random tie-breaking rule.

To consider this interpretation formally, suppose that in the event of equal bids, instead of sharing the supplies equally, one and only one of the merchants with the high bid is chosen at random, say with probability $\alpha \in (0,1)$.[8] This winner gets all the stock and is a monopolist in the second stage. Readers can verify that Lemma 2 continues to hold, so the only possible SPNE strategy is the zero-monopoly-rent bid $\tilde{p}$. However, now the second-stage outcome in the event both merchants bid $\tilde{p}$ is not a Bertrand game, but instead is a monopoly with second-stage revenue $M[S(\tilde{p})]$. Expected profits are $E\pi = \alpha\mu(\tilde{p}) = 0$, as anticipated by the merchants. Therefore, $\tilde{p}$ is an SPNE bid, even for inelastic demand (i.e., when $p_c < \hat{p}$).

---

[6] The sum of the payoff functions, $E\pi_1 + E\pi_2$, is not upper semicontinuous, thus not satisfying the conditions for Theorem 5 of Dasgupta and Maskin (in particular, see their Example 3).

[7] To illustrate this non-intuitive phenomenon, let F_{in} for $i = 1,2$ be independent identical uniform probability distributions on the finite set $(1/n^2, 2/n^2, \ldots, 1/n)$. As $n \to \infty$, F_{in} converge weakly to the degenerate probability distribution with unit mass at 0, because for all $\varepsilon > 0$ there is an n_ε such that for all $n > n_\varepsilon$ the F_{in} probability of the interval $[\varepsilon, \infty)$ is precisely zero. Furthermore, the probability of a tie (i.e., identical draws, one from each distribution) is $1/n$, which obviously goes to zero as $n \to \infty$.

[8] The discrete price version suggests one way of approximating this scheme without explicitly imposing it. For an analogous R&D game, D. Fudenberg and J. Tirole (1985) provide a formal way to achieve tacit random tie breaking. If the first stage of our merchant game were organized as a descending bid auction, then their results would imply that each merchant would have a 50:50 chance of winning.

PROPOSITION 3: *Under a random tie-breaking rule, there exists a unique SPNE, in which* (1) $p_{bi} = \tilde{p}$ *the zero-monopoly-rent bid for both merchants, and* (2) *the randomly selected high bidder is a monopolist in the second stage.*

In the low elasticity case ($\eta_c < 1$), this SPNE outcome is *non-Walrasian and inefficient*. The bid and ask prices exceed the Walrasian price: $\hat{p} > \tilde{p} > p_c$, and since the winner has excess stock $x = S(\tilde{p}) > D(\hat{p})$, he throws away the excess. (To see this, note that the point $(\tilde{p}, \hat{p})$ lies to the right of the $g(\;)$ locus in Figure 2A.) In the high elasticity case ($\eta_c \geq 1$), the outcome is Walrasian (and efficient) since in that case the zero-monopoly-rent bid $\tilde{p} = p_c$.

When demand is sufficiently elastic ($\eta_c \geq 1$), while both versions of the model (the equal-sharing rule and the random tie-breaking rule) yield the Walrasian price as the unique SPNE price, they are very different in terms of the number of active merchants in the second stage. Under the equal-sharing rule, all merchants are active, whereas under the random tie-breaking rule, the consumers face a monopolist. The fact that the equal-sharing rule has all merchants active in both stages is an appealing realistic result. For this reason, in my mind the better model choice is the equal-sharing rule when that model has an SPNE (i.e., for elastic demand), and the random tie-breaking rule when demand is inelastic, with random tie breaking interpreted as the "proper" limit of discrete price approximations of the equal-sharing version (as discussed above).

It is noteworthy that the SPNE outcome of all models discussed leave both merchants with zero profits. This is reminiscent of Demsetz's (1968) conjecture that competition for a monopoly license will drive rents to zero. (This is the motivation for the label "zero-monopoly-rent" bid.)

The specifications of Section I rule out the cases of vertical and horizontal supply. The case of vertical supply goes through provided that $D(0)$ exceeds the fixed supply, which ensures that $p_c > 0$. This condition avoids the discontinuity in supply at a bid price of 0. When this condition is not fulfilled, we essentially have a horizontal supply at zero marginal costs. The case of horizontal supply defines a distinctly different game, since the bidding subgame no longer has the winner-take-all characteristic. In particular, we get the Kreps and Scheinkman "capacity choice" game in which there is no competition for inputs and the outcome is Cournot.[9] Thus, unrestricted availability of inputs at a constant marginal cost is a singular case.

IV. Duopolist Merchants with Forward Contracts

Reversing the stages of the previous game results in a model of merchants who first sell forward contracts to consumers and then obtain supplies to fill these contracts. Bertrand price competition is assumed in both stages. In addition, the penalty for default is assumed to be sufficiently severe to

---

Another remedy of the nonexistence problem would be to allow the merchants to trade among themselves after the first stage. The resulting SPNE would have one merchant purchasing all the supplies of the others at cost; thus, the outcome would look like random tie breaking.

[9] This result may appear at odds with X. Vives (1986) who shows that flexible technology (a flat marginal cost curve) leads to p_c as the SPNE outcome. However, the games are very different. In Vives, when the marginal cost curve is flat, the first-stage capacity choice is superfluous since the firms can produce any additional amount at a fixed marginal cost. In contrast, in our merchant model the first-stage quantity choice (via bids) is binding as in Kreps and Scheinkman.

prevent any default in equilibrium. A penalty that gives a defaulting merchant a nonpositive payoff will suffice. This assumption also rules out threats off the equilibrium path involving default.

In the case of default, I assume that the legal proceedings are sufficiently lengthy that the unsatisfied consumers are out of the market until some later period. Thus, merchants cannot gain additional customers by forcing another merchant into default.

Obviously, the monopolist solution to this forward contract model is exactly the same as described in Section II (with the order reversed). The duopolist solution, however, is different from Section III.

The first stage is now the consumer market. Since there are no capacity constraints in forward contracts, the lowest ask price gets all the orders; there is no need for a rationing mechanism when ask prices differ; consequently, we have no need to invoke the RDC for this model. In the event of equal ask prices, assume (for now) that orders are split equally. At this point the intuition is clear: since there are no essential capacity constraints, Bertrand competition will yield Walrasian prices uniquely.

To derive this result rigorously in this two-stage game, let $y_i(p_{a1}, p_{a2})$ denote the realized orders in stage one.

$$y_1(p_{a1}, p_{a2}) = \begin{cases} D(p_{a1}), & \text{if } p_{a1} < p_{a2} \\ D(p_{a1})/2, & \text{if } p_{a1} = p_{a2} \\ 0, & \text{if } p_{a1} > p_{a2}. \end{cases}$$

As in Section III, let $x_i(p_{b1}, p_{b2})$ denote the realized stock, now from stage two. The no-default condition requires that $y_i(p_{a1}, p_{a2}) \leq x_i(p_{b1}, p_{b2})$. Subgame perfection rules out noncredible threats. In particular, if $p_{a1} > p_{a2}$, then merchant 1 gets no orders, so he would not want any supplies; hence, his bid should be strictly less than merchant 2's bid $[p_{b1} < p_{b2}]$. Conditional on the first-stage play, no-default and subgame perfection together put the following requirement on second-stage bids of merchant 1 (and symmetrically for merchant 2): (1) If $p_{a1} < p_{a2}$, then $p_{b1} \geq S^{-1}[D(p_{a1})$. (2) If $p_{a1} = p_{a2}$, then $p_{b1} \geq \max\{S^{-1}[D(p_{a1})], p_{b2}\}$.

Condition (*i*) says that when a merchant wins the first stage subgame, then he must bid sufficiently high to secure the necessary supplies to fill his orders.

Condition (*ii*) says that when both merchants share the orders, each must bid sufficiently high to fill his orders.

Define strategy σ_i as follows: (*i*) $p_{ai} = p_c$; and (*ii*) $p_{bi} = S^{-1}[D(p_{ai})]$ if $p_{ai} \leq p_{aj}$, and $p_{bi} \in [0, S^{-1}[D(p_{aj})])$ otherwise. To elucidate (2), note that when $p_{a1} < p_{a2}$, merchant 1 gets all the orders and bids $p_{b1} = S^{-1}[D(p_{a1})]$; then merchant 2 can bid anything from 0 up to but not including p_{b1} all of which gives him no supplies. This leeway in the losing bid implies a multiplicity of equilibrium strategies; but this is the only non-uniqueness. The *Walrasian outcome*, defined as $p_{ai} = p_c = p_{bi}$ for all i, is unique.

PROPOSITION 4: *The family of symmetric strategies* $\{\sigma_1, \sigma_2\}$ *coincides with the set of SPNE strategies, all of which yield the Walrasian outcome.*

PROOF:

(1) σ yields payoffs of $E\pi = 0$. If $p_{a1} > p_c$, merchant 1 obtains no orders and underbids, yielding a zero payoff, which is no better. If $p_{a1} < p_c$, then merchant 1 gets all the orders, and must bid at least $S^{-1}[D(p_{a1})]$ which is greater than p_c; hence, $E\pi < 0$. Therefore, σ constitutes an NE. (2) σ is subgame perfect because the specified p_{bi} are NE of the second-stage subgame. (3) The uniqueness proof is given in the Appendix.

Thus, we see that the Bertrand merchant game with forward contracts yields the Walrasian price as the unique SPNE outcome regardless of the relation of p_c and the sales-revenue-maximizing price $\hat{p}$ (or the elasticity of demand η_c). Note also that all merchants are active in both stages.

Proposition 4 extends to games with any number of merchants greater than one. The proof mimics that given here. Hence, "*two*" *is sufficient for perfect competition*. It should go without saying that while two is sufficient for perfect competition in this abstract model, we are certainly not suggesting that two is sufficient in reality. Numerous "realistic" modifications (for example, differentiation among the merchants) would affect the results.

This result is also robust to the allocation scheme in the event of equal prices as long as both stages are coordinated; that is, in the event of equal ask prices if merchant 1 gets a fraction α of the orders, then in the event of equal bids he must get α of the supplies. Coordination is needed to avoid defaults induced by the allocation scheme.

The extension of these results to the cases of vertical and horizontal supply are also different from Section III. When supply is vertical at $S < D(0)$, the SPNE has $p_{ai} = p_c$ for both players, but now we have a continuum of SPNE associated with the multiplicity of second-stage NE with both merchants bidding a common price between 0 and p_c. Essentially we have an indeterminacy in who gets the surplus—the merchants or the suppliers. When $S \geq D(0)$, we are really in the case of horizontal supply, and Proposition 4 holds. Of course, if supply is horizontal at p', then $p_c = p'$.

V. Discussion

The notion of "market-making" merchants who buy from suppliers and sell to consumers has been modeled as a one-shot, two-stage Bertrand game. With one exception, the SPNE prices are the Walrasian prices. This result holds for any number of merchants greater than one. Thus, we have a (albeit partial equilibrium) model in which the Walrasian price arises not from the benevolent actions of a fictitious auctioneer but from optimal price-setting behavior of merchants. A robust characteristic of the SPNE (when it exists) with or without forward contracts is zero merchant profits.

These results stand in sharp contrast to the results for capacity-constrained Bertrand competition even with capacity choice. The crucial difference is that in this paper there is "winner-take-all" competition for the inputs, whereas in the other papers either the inputs are taken as fixed or each player has an independent source for capacity production. The lesson is that it is not safe to ignore the input side or simply assume unrestricted availability of inputs at a constant marginal cost. Indeed, the difference between the Cournot outcome (under the latter assumption) and the zero-profit outcome indicates an incentive for vertical integration.

Another lesson is that whether the first move is with quantities or with prices is crucial. Indeed, the first move is all important. When quantity choice precedes price competition (as in Kreps and Scheinkman, 1983), the outcome is Cournot, so price competition is ineffectual (i.e., it fails to produce Walrasian price). However, when price competition precedes quantity choice (as in the forward contract version of this paper), the outcome is Walrasian so quantity choice is ineffectual (i.e., it fails to support Cournot outcomes or collusive outcomes).

It is surprising that without forward contracts and the equal-sharing rule for tie bids, an SPNE does not exist when demand is inelastic ($\eta_c < 1$). On the other hand, under a random tie-breaking rule, a unique SPNE exists, and I have argued that this SPNE is the proper limit of finite approximations under the equal-sharing rule.

When the elasticity of demand is at least unity, both the model with forward contracting and the model with supplies first and equal sharing yield Walrasian prices and all merchants active in both stages. However, when the elasticity of demand is less than one, the outcomes differ substantially. The forward contracting model yields the Walrasian outcome with all merchants active in both stages, whereas the model with supplies first and random tie breaking yields non-Walrasian prices, a second-stage monopolist, and waste. This is an empirically observable difference. (I leave the examination of the empirical evidence to those who have a comparative advantage in that.)

In the models considered here, when demand is inelastic, consumers are better off with forward contracting, while suppliers are better off without forward contracting (provided random tie breaking is effected). How these conflicting interests get resolved in the choice of a market institution is a topic for future research.

While the paper focused on a single play of the two-stage game, since the SPNE outcomes are unique, the same strategies will constitute the unique SPNE of every finitely repeated version of the game, and the outcome will be Walrasian in every period (with the one exception noted).[10] A crucial assumption in this extension to repeated play is that inventories not accumulate from one period to the next. The effect of inventory accumulation is the proper subject of future research.

APPENDIX

PROOF OF LEMMA 1

(1) The sales revenue (i.e., payoff) functions of pure-strategy ask prices, $R_i(p_{a1}, p_{a2}) = p_{ai}z_i(p_{a1}, p_{a2}; x)$, are continuous except possibly when $p_{a1} = p_{a2}$. It is straightforward to verify that $R_1 + R_2$ is upper semicontinuous given RDC (1) and (2), and that R_i is weakly lower semicontinuous. Therefore, by Theorem 5 of Dasgupta and Maskin, an NE exists.

(2) Let (Ψ_1, Ψ_2) denote a pair of (possibly mixed) NE strategies. Let $\underline{p}_i$ denote the lower bound of the support of Ψ_i. I claim that $\underline{p}_1 = \underline{p}_2 = \underline{p} \geq D^{-1}(2x)$. (a) Suppose to the contrary that $\underline{p}_1 > \underline{p}_2$. Then for all $p_{a2} < \underline{p}_1$, the expected payoff to merchant 2 of playing pure-strategy p_{a2} against Ψ_1 is $ER_2(p_{a2},\Psi_1) = p_{a2}\min(x, D(p_{a2}))\}$. Since Ψ_2 is a best response, $\underline{p}_2$ must be a local maximum of $ER_2(\cdot,\Psi_1)$, which implies that $\underline{p}_2 \geq \hat{p}$. Then, given the RDC, it follows that $ER_2(\cdot,\Psi_1)$ is strictly decreasing for all $p_{a2} > \underline{p}_2$, so Ψ_2 must be concentrated at $\underline{p}_2$. But then, $ER_1(p_{a1},\Psi_2) = 0$ for all p_{a1} in the support of Ψ_1. In contrast, $ER_1(p_{a1} = \underline{p}_2, \Psi_2) = \underline{p}_2 D(\underline{p}_2)/2 > 0$. Therefore, Ψ_1 is not a best response. A symmetric argument rules out $\underline{p}_2 > \underline{p}_1$; thus, there exists a $\underline{p}$ such that $\underline{p}_1 = \underline{p}_2 = \underline{p}$ as claimed. (b) Finally, suppose $\underline{p} < D^{-1}(2x)$. From RDC (2),

$$\lim_{p_{a1} \to \underline{p}} \mathrm{RD}(p_{a1}, \underline{p}; x) = \max\{D(\underline{p}) - x, 0\} > x.$$

Thus, for small $\varepsilon > 0$, $p_{a1} = \underline{p} + \varepsilon$ would yield discretely more sales (and profits), so Ψ_1 could not have been a best response. Thus, $\underline{p} \geq D^{-1}(2x)$ as claimed.

(3) If $\underline{p} > D^{-1}(2x)$, then $\Psi_i\{\underline{p}\} = 0$. This follows because by undercutting slightly one merchant can discretely increase his sales from $D(\underline{p})/2$ to $\min\{D(\underline{p} - \varepsilon), x\}$.

(4) I now claim that, given x, both merchants receive the same equilibrium expected payoff: $ER_i(x) = ER(x)$ for both merchants. First, note that all p_i in the support of Ψ_i must yield the same expected profits, so $ER_1 = ER_1(\underline{p}, \Psi_2)$, the expected revenues from playing $\underline{p}$ against Ψ_2. There are two cases. (a) If $\underline{p} = D^{-1}(2x)$, then $ER_i = \underline{p}x$ for both merchants. (b) If $\underline{p} > D^{-1}(2x)$, then from step (3) above, $ER_i = \underline{p}\min\{x, D(\underline{p})\}$ for both merchants. Thus, in both cases, ER_i can be expressed as a function $ER(x)$.

(5) Let P_i denote the upper bound of the support of Ψ_i. Then, we must have $ER_1(P_1, \Psi_2) = ER_2(P_2, \Psi_1) = ER(x)$. Without loss of generality, suppose $P_1 \geq P_2$. Then $ER(x) = ER_1(P_1, \Psi_2) =$

$$P_1 \int_{y < P_1} \min\{x, \mathrm{RD}(P_1, y)\}\, d\Psi_2 + P_1 \min\{x, D(P_1)/2\}\Psi_2\{P_1\},$$

which by RDC(3) is $\leq P_1\min\{x, D(P_1)/2\}$. But then since $p\min\{2x, D(p)\} \leq M(2x)$, it follows $ER_i = ER(x) \leq M(2x)/2$.

PROOF OF LEMMA 2

To obtain the most general result, we must allow for the possibility of mixed bid strategies as well as pure bid strategies. Let F_i denote a cumulative probability distribution which characterizes merchant i's first stage (possibly mixed) strategy for p_{bi}. Recall that the support of F_i is the smallest closed set with F_i probability one. Let $\underline{p}_i$ denote the lower bound of the support of F_i. Since clearly the support of F_i cannot contain bid prices above $\tilde{p}$, it will suffice to prove that $\underline{p}_i = \tilde{p}$ for all i. First note that if $\underline{p}_1 < \tilde{p}$, then $\underline{p}_2 < \tilde{p}$, because merchant 2 would get $E\pi_2 \leq 0$ by choosing $\underline{p}_2 \geq \tilde{p}$, whereas she could get $E\pi_2 > 0$ by choosing $\underline{p}_2$ slightly below $\tilde{p}$. Second, observe that we cannot have both $\underline{p}_i < \tilde{p}$. If $\underline{p}_1 < \underline{p}_2 < \tilde{p}$, then all $p_{b1} < \underline{p}_2$ are losing bids yielding zero payoff, whereas there are bids $p_{b1} \geq \underline{p}_2$ that would yield positive payoffs with positive probability. Therefore, F_1 is not a best response to F_2; and symmetrically for $\underline{p}_2 < \underline{p}_1 < \tilde{p}$. If $\underline{p}_1 = \underline{p}_2 < \tilde{p}$, then using Lemma 1, for small $\varepsilon > 0$, $p_{b2} = \underline{p}_2 + \varepsilon$ would yield strictly greater expected profits, so F_2 could not be a best response. Therefore, $\underline{p}_i = \tilde{p}$ for all i.

---

[10] Of course, the Folk theorems (for example, Benoit and Krishna, 1985; Fudenberg and Maskin, 1986) tell us that, in addition to the "noncooperative" SPNE, collusive outcomes can be sustained as SPNE of infinitely repeated games. However, we do not have a theory of how a particular outcome arises from the multiplicity of possible SPNE. In any event, the vital ingredient of trigger strategy supergame equilibria (J. Friedman, 1971) is the equilibrium of the one-shot game, so this paper is a necessary starting point.

Extending Propositions 1 and 2 to $N > 2$. The key ingredient was Lemma 1. One can show analogously that for any $N \geq 2$, $ER(x) \leq M(Nx)/N$. Then $ER(x) - xS^{-1}(Nx) \leq [M(Nx) - NxS^{-1}(Nx)]/N$, so the overbidding argument goes through. The rest is straightforward.

Proof of Uniqueness Part of Proposition 4. Let G_i denote the SPNE (possibly mixed) strategies for p_{ai}, and let P_i denote the upper bound of supp$\,G_i$. Clearly, $P_i \geq P_c$, since otherwise $E\pi_i < 0$. Note that if $P_1 > p_c$, then $P_2 > p_c$. To see this, suppose to the contrary that $P_2 = p_c$, so $E\pi_2 \leq 0$. But by choosing a pure strategy $p_c + \varepsilon$ (for some $\varepsilon > 0$), she wins all the orders with some positive probability, so $E\pi_2 > 0$, which proves the claim. Now observe that we cannot have both P_i above p_c. If $P_1 > P_2 > p_c$, then G_1 could not possibly be a best response; and symmetrically for $P_2 > P_1 > p_c$. If $P_2 = P_1 > p_c$, then for small $\varepsilon > 0$, $p_{a2} = P_2 - \varepsilon$ will yield strictly greater expected profits. Therefore, $P_1 = p_c$ for all i.

REFERENCES

Allen, B. and Hellwig, M., "The Range of Equilibrium Prices in Bertrand-Edgeworth Duopoly," Center for Analytic Research in Economics and the Social Sciences, Working Paper No. 85-01, University of Pennsylvania, 1984.

______ **and** ______, "Price-Setting Firms and Oligopolistic Foundations of Perfect Competition," *American Economic Review Proceedings*, May 1986, *76*, 387–92.

Beckman, M. J., "Edgeworth-Duopoly Revisited," in R. Henn, ed., *Operations Research — Verfahren, III*, 1965: Verlag Anton Hein, 55–68.

Benassy, J-P, "On Competitive Market Mechanisms," *Econometrica*, January 1986, *54*, 95–108.

Benoit, J-P and Krishna, V., "Dynamic Duopoly; Prices and Quantities," *Review of Economic Studies*, January 1987, *54*, 23–35.

Brock, W. and Scheinkman, J., "Price-Setting Supergames with Capacity Constraints," *Review of Economic Studies*, July 1985, *52*, 371–82.

Dasgupta, P. and Maskin, E., "The Existence of Equilibrium in Discontinuous Economic Games: Theory and Applications," *Review of Economic Studies*, January 1986, *53*, 1–42.

Davidson, C. and Deneckere, R., "Long-Run Competition in Capacity, Short-Run Competition in Prices, and the Cournot Model," *Rand Journal of Economics*, Autumn 1986, *17*, 404–15.

Demsetz, H., "Why Regulate Utilities?," *Journal of Law and Economics*, April 1968, *11*, 55–66.

Dixon, H., "Existence of Mixed Strategy Equilibria in a Price-Setting Oligopoly with Convex Costs," *Economic Letters*, 1984, *16*, 205–12.

Dubey, P., "Price-Quantity Strategic Market Games," *Econometrica*, May 1982, *50*, 111–26.

Edgeworth, F., "The Pure Theory of Monopoly," Papers Relating to Political Economy, London: Royal Economic Society, 1925, 111–42.

Friedman, J., "A Noncooperative Equilibrium for Supergames," *Review of Economic Studies*, January 1971, *38*, 1–12.

Fudenberg, D. and Maskin, E., "The Folk Theorem for Games with Discounting and Incomplete Information," *Econometrica*, May 1986, *54*, 533–54.

______ **and Tirole, J.**, "Preemption and Rent Equalization in the Adoption of New Technology," *Review of Economic Studies*, May 1985, *52*, 383–401.

Kreps, D. and Scheinkman, J., "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics*, Autumn 1983, *14*, 326–38.

Levitan, R. and Shubik, M., "Price Duopoly and Capacity Constraints," *International Economic Review*, February 1972, *13*, 111–21.

Maskin, E., "The Existence of Equilibrium with Price-Setting Firms," *American Economic Review Proceedings*, May 1986, *76*, 382–86.

Osborne, M. and Pitchik, C., "Price Competition in a Capacity-Constrained Duopoly," *Journal of Economic Theory*, April 1986, *38*, 238–60.

Simon, L., "Bertrand, the Cournot Paradigm and the Theory of Perfect Competition," *Review of Economic Studies*, April 1984, *51*, 209–30.

Vives, X., "Commitment, Flexibility, and Market Outcomes," *International Journal of Industrial Organization*, June 1986, *4*, 217–29.

On The Interpretation of Near Random-Walk Behavior in GNP

By KENNETH D. WEST*

Several recent papers have studied the univariate time-series process for U.S. GNP, including John Campbell and Gregory Mankiw, 1986; Peter Clark, 1986a, b; John Cochrane, 1986; Charles Nelson and Charles Plosser, 1982; Danny Quah, 1986; James Stock and Mark Watson, 1986; and Watson, 1986. A major focus of these papers has been the extent to which GNP movements are well approximated by a process with a unit root with drift, as opposed to stationary movements around a time trend. The empirical evidence on this is mixed. Campbell and Mankiw, 1986; Nelson and Plosser, 1982; and Stock and Watson, 1986, conclude that the random-walk (unit-root) approximation is quite good. Clark, 1986a, b; Cochrane, 1986, and, perhaps, Quah, 1986; and Watson, 1986, say that it is not.

Campbell and Mankiw, 1986, and Nelson and Plosser, 1982, both argue that if the random-walk approximation in fact is reasonable, there are important implications for business cycle theory. This is because movements in random walks are permanent: a shock today has an infinitely long-lived effect. The concept of a stationary natural rate, Campbell and Mankiw note, has little utility if a GNP shock is, on average, never offset by a return to some trend rate of GNP. Nelson and Plosser suggest that monetary disturbances are unlikely to be an important source of GNP fluctuations, since monetary shocks are typically thought to have no permanent effect. Both conclude that if the random-walk characterization is accurate, an implication is that fluctuations in GNP are unlikely to be driven by nominal demand shocks.[1] Similar inferences appear to be drawn by Rene Stultz and Walter Wasserfallen, 1985; Angus Deaton, 1986; and Olivier Blanchard and Quah, 1987.

Campbell and Mankiw, 1986, and Nelson and Plosser, 1982, of course recognize that their random-walk characterization is only a convenient approximation. In any finite sample, it will not be possible to discriminate between a unit root (random walk) and a root arbitrarily near, but below, unity (what this paper calls a "near random walk"). This is potentially a practical problem. The Monte Carlo evidence in David Dickey and Wayne Fuller, 1981, indicates that with Nelson and Plosser's 1982 sample size (less than 100), Nelson and Plosser's test of a unit-root null is not very likely to reject even when the true process is stationary, with autoregressive coefficients whose sum is as low as .8. Coefficients of this size and larger are suggested by studies that assume the GNP process is stationary. An AR(2) of log real GNP around trend fitted to annual U.S. data 1948–1985, for example, yields coefficients whose sum is .83; since the estimate of this sum is sharply downward biased for processes with near unit roots (Fuller, 1976), the .83 point estimate is suggestive of a sum even closer to unity.[2]

---

*Woodrow Wilson School, Princeton University, Princeton, NJ 08544. I thank Alan Blinder, Ben Bernanke, Mark Gertler, Robert J. Gordon, Bennett T. McCallum, and an anonymous referee for helpful comments and discussions, and the National Science Foundation for financial support.

[1] Nelson and Plosser, 1982, p. 166, conclude that "assigning a major portion of variance in output to the innovation in [a] nonstationary component gives an important role for real factors in output fluctuations and places limits on the importance of monetary theories of the business cycle." Campbell and Mankiw, 1986, p. 24, state that their results are "inconsistent with many prominent theories in which output fluctuations are primarily caused by shocks to aggregate demand...[including] models based on long-term nominal contracts."

[2] Throughout, I assume annual rather than quarterly data, for two reasons. The first is for consistency with some of the relevant studies, including John Taylor, 1980b, and Nelson and Plosser, 1982. The second is that the two-period contract length that, for simplicity, will

The aims of this paper are twofold. The first is to point out that it is dangerous to use a single country's univariate GNP process to draw structural inferences concerning the stability of the natural rate, or of the importance of nominal shocks in business cycles, given that in practice one cannot discriminate between random walk and near random-walk behavior. The second is to emphasize that simple natural rate models with nominal shocks are as capable as simple real business cycle models (for example, Robert King et al., 1987) in generating a highly persistent process for GNP.

The paper uses a variant of John Taylor's (1980a, b) overlapping wage contracts model, which maintains a stationary natural rate.[3] In my variant (unlike Taylor's) the only source of instability—the only reason GNP ever deviates from the natural rate—is shocks to monetary policy. Thus, monetary policy is the only important factor in the business cycle. It is shown that near random-walk behavior in GNP can result from monetary policy of the sort often attributed to the U.S. Federal Reserve.

The basic idea is as follows. In practically any model, including Taylor's, serial correlation in movements of the money stock puts serial correlation in movements in prices. In Taylor's model, prices do not adjust instantaneously to movements in money. Additional persistence in prices is induced by the overlapping wage contracts. Movements in real interest rates and real balances therefore are serially correlated, and this induces serial correlation in aggregate demand and GNP. The degree of the serial correlation depends on the monetary authority's money supply rule and the model's basic parameters. Stylized versions of simple money supply rules, and plausible values for the model's basic parameters, suggest near random-walk behavior in GNP. The implied autoregressive root is about .8 to .99.

Near random-walk behavior, then, is perfectly consistent with Taylor's natural rate model. This is, of course, implicitly a message in Taylor (1980a, b), since it is argued there that, at least in the presence of supply shocks, the model is capable of tracking observed movements in GNP. The present paper generalizes Taylor's result in two ways. First, I show that near random-walk behavior results even in a model with purely nominal shocks. In light of Campbell and Mankiw's 1986, and Nelson and Plosser's 1982 interpretation of their results, this seems important to establish. Second, I show that near random-walk behavior results even in a version of Taylor's model extended to include standard IS and LM curves, with a monetary policy rule of targeting the interest rate. Given the widespread use of such an aggregate demand apparatus (at least in textbooks), this seems to be an useful generalization.

Before deriving my results, let me emphasize what I am *not* arguing. I am not arguing that destablizing monetary policy is the sole, or even most important, source of U.S. GNP's near random-walk behavior. I am suppressing the supply and demand shocks present in Taylor (1980a, b) not because I doubt their importance, but to make my point as cleanly and emphatically as possible. I am also not arguing that the unit-root approximation is a bad one. It is probably appropriate for forming simple ARIMA forecasts (Bennett McCallum, 1986a), for example. It may even be appropriate for structural estimation and inference: I am not arguing for a stationary monetary theory of the business cycle against, say, a nonstationary real theory. My point, rather, is that the stylized facts about GNP are perfectly consistent with Taylor's widely used model. Simple analysis of a given country's univariate process for GNP therefore is unlikely to be particularly helpful in distinguishing stationary from nonstationary theories, or between models in which nominal shocks are very important from those in which they have negligible effects.

---

be assumed in Section I below, is implausibly short for quarterly but not for annual data. In addition, and again for consistency, all empirical estimates are taken from studies using post-World War II U.S. data.

[3] I follow Taylor (1980a, b) in interpreting his model as a natural rate one. Bennett McCallum, undated, 1987, argues otherwise.

I. Near Random Walk for GNP

The aggregate supply curve (Phillips curve) is borrowed from Taylor (1980a, b). There are staggered two-period wage contracts. In each period, one half of the labor force fixes its nominal wage for the next two periods.

$$(1) \qquad x_t = .5x_{t-1} + .5_{t-1}x_{t+1} + .5\gamma({}_{t-1}y_t + {}_{t-1}y_{t+1}),$$

$$(2) \qquad p_t = .5(x_t + x_{t-1}),$$

$$(3) \qquad y_t = -\delta(i_t - {}_tp_{t+1} + p_t),$$

$$(4) \qquad m_t - p_t = \theta_1 y_t - \theta_2 i_t,$$

$$(5) \qquad i_t = \lambda i_{t-1} + u_t.$$

The variables are x_t = log nominal contract wage, y_t = log GNP, i_t = nominal interest rate, p_t = log price level, m_t = log money supply, u_t = a serially uncorrelated shock. A "$t-1$" subscript, as a prefix, denotes expectations at time $t-1$. All variables are zero-mean deviations from trend. Trend GNP is by definition potential or natural rate GNP.

Equation (1) says that the nominal wage depends on actual and expected wages, as well as expected demand pressure. The latter is measured by expected deviations of GNP from trend. Equation (2) is a price markup equation. Equation (3) is a standard IS curve, relating GNP to the *ex ante* real interest rate. Equation (4) is a standard LM curve, expressing the demand for money as a function of the nominal interest rate and GNP. As noted in the introduction, the supply and demand shocks that quite plausibly are present in equations (1) to (4) are suppressed, to emphasize the potential role for monetary policy in output fluctuations.

Equation (5) is the money supply rule, with $0 \leq \lambda < 1$ and u_t a serially uncorrelated shock. The monetary authority is thus assumed to smooth movements in interest rates. Empirical evidence that i_t followed a near random walk in the postwar period (λ is near one) may be found in Eugene Fama and Michael Gibbons (1982). A theoretical argument why the Fed might have set nominal interest rates to follow a near random walk may be found in Mankiw (1986).

It is straightforward, though tedious, to solve the model.[4] Let

$$(6) \quad a = -(1+\gamma^{-1}\delta^{-1}) + \gamma^{-1}\delta^{-1}(4\gamma\delta+1)^{1/2},$$
$$b = 2\gamma\delta(\lambda+\lambda^2)/\{(\lambda-1)(a\gamma\delta+\lambda\gamma\delta+2+2\gamma\delta)\}.$$

The Appendix shows that the contract wage x_t obeys

$$(7) \qquad (1-aL)(1-\lambda L)x_t = bu_t,$$

where L is the lag operator. In conjunction with the price markup equation (2), equation (7) can be used to solve for the stochastic process for p_t. When this is plugged into the IS curve (3), one can calculate the stochastic process for y_t. Output in general follows an ARMA (2,2) process,

$$(8) \quad y_t = [\eta(L)/\rho(L)]u_t,$$
$$\rho(L) \equiv (1-aL)(1-\lambda L),$$
$$\eta(L) \equiv \delta[(-1+.5b)+aL-.5bL^2]$$
$$= (-\delta+.5b\delta)\left[1+(-1+.5b)^{-1}aL - (-1+.5b)^{-1}(.5b)L^2\right]$$
$$\equiv d_0(1+d_1L+d_2L^2).$$

To see how the properties of (8), the univariate process y_t, depend on the monetary policy rule (depend on λ), consider two cases. The first is $\lambda = 0$, $i_t = u_t$. Since $\lambda = 0$ implies $b = 0$ (see equation (6)), we have from (8) that

---

[4] The well-known indeterminacy of rational expectations models under interest rate rules (Thomas Sargent, 1979; McCallum, 1986b), applies here as well. The rule (5) is interpreted as in McCallum (1986b) as the limit of a certain non-interest rate rule that yields a unique stationary solution for y_t. The restriction $\lambda \neq 1$ is imposed because for $\lambda = 1$ this solution technique breaks down (a divide by zero is implied). See the formula for $\underline{b}$ in equation (6) below.

$$y_t = -\delta u_t. \tag{9}$$

So if the monetary authority takes care that the current nominal interest rate is independent of past shocks, deviations of output from the natural rate are serially uncorrelated.

Consider instead the case when λ is near unity. From (6), as $\lambda \to 1$, $b \to -\infty$. It follows that for λ arbitrarily near one, d_1 (defined in (8)) will be arbitrarily near zero and d_2 arbitrarily near -1. Thus for λ very near 1, $\eta(L)$ will factor as $(-\delta + .5b\delta)(1+\eta_1 L)(1-\eta_2 L)$, $\eta_1 \approx 1$, $\eta_2 \approx 1 \approx \lambda$. Since the $1-\eta_2 L$ factor approximately cancels the $1-\lambda L$ autoregressive factor, y_t will behave very much like the ARMA(1,1) process that results when these factors are canceled,[5]

$$y_t \approx d_0[(1+L)/(1-aL)]u_t. \tag{10}$$

It follows that y_t will behave much like a variable with a unit root if $\underline{a}$ is near one.

We have $a \to 1$ as $\gamma\delta \to 0$, that is, as the aggregate supply and/or IS curves become horizontal. Taylor (1980b) estimated γ to be about .05 to .10. Jeffrey Sachs's (1980) Phillips curve regressions suggest that γ is about .01 to .07.[6] Benjamin Friedman's (1977) estimates suggest $\delta = .17$;[7] Taylor (1985) indicates that δ is less than .125. If we take .01 to .10 as the range for γ, .1 to .2 as the range for δ, then $\gamma\delta$ is about .001 to .02. This yields a range for $\underline{a}$ of about .96 to .998. See Table 1, Part A. Plausible parameter values therefore suggest that the near random-walk characterization will be quite good if the monetary authority attempts to stabilize interest rates by setting ${}_{t-1}(i_t - \lambda i_{t-1})$ to zero for λ near one. With $\underline{a}$ this near unity, it will be difficult to reject the null hypothesis of a unit root, in sample sizes typically available.[8]

TABLE 1—IMPLIED AUTOREGRESSIVE ROOT FOR GNP

A. Interest Rate Target	
Range for structural parameters:	
sensitivity of wages to excess demand (γ):	.01 to .10
slope of IS curve (δ):	.1 to .2
Range for implied AR root:	.96 to .998
B. Money Supply Target	
Range for structural parameters:	
sensitivity of wages to excess demand (γ):	.01 to .10
degree of monetary accommodation (g):	.3
Range for implied AR root:	.78 to .93

Note: The model for panel A consists of equations (1) to (5); for panel B of equations (1), (2), (11), and (12). In each case λ is assumed at or near unity.

The intuition to the effect of λ on the univariate y_t process is as follows. With $\lambda = 0$, the contract wage and price level are nonstochastic: $x_t = x_{t-1} = 0$ is the only stationary (constant mean) solution to (7) with $b = 0$. So the IS curve (3) implies $y_t = -\delta i_t$, and, with $\lambda = 0$, i_t is serially uncorrelated. By contrast, when $\lambda \neq 0$, the autoregressive root of $1-\lambda L$ in the monetary authority's control variable puts the same root in the wage and price processes; the long-run properties of the money supply rule of course are reflected in wages and prices. But that is not all. As Taylor (1980a) has emphasized, overlapping contracts can be an endogenous source of persistence. The serial correlation in the money supply induces serial correlation in wages and prices above and beyond that directly produced by the $1-\lambda L$ root. So expected inflation, ${}_tp_{t+1} - p_t$, does not move instantaneously, and one to one, with i_t. The real interest rate is serially correlated, and, therefore, as per the IS curve (3), so is GNP.

---

[5] This illustrates the possibility that approximate cancellation of common autoregressive and moving average factors may help explain Andrew Rose's (1987) result that univariate time-series have simpler ARMA representations than are suggested by multiequation structural models. See Rose, 1987, pp. 27–29.

[6] Rewrite (1) as $x_t - x_{t-1} = {}_{t-1}(x_{t+1} - x_t) + \alpha y_t + e_t$, $\alpha \equiv 2\gamma$, $e_t \equiv -(x_t - {}_{t-1}x_t) - \gamma(y_t - {}_{t-1}y_t + y_t - {}_{t-1}y_{t+1})$. This is in the usual Phillips curve form, inflation = expected inflation + α * excess demand + shock. The .01 to .07 range reported in the text is one half of Sachs's (1980) estimates of α (that is) one half of his post-war estimates of the coefficients that he calls β_1 and ϕ in his Tables 3 and 4).

[7] This is Friedman's (1977, p. 322) implied estimate of the long-run elasticity of real spending with respect to the nominal interest rate, from regressions using quarterly data. [The short-run (single quarter) elasticity is .09.] Friedman (1977, p. 323) notes that in his 1961–76 sample period, nominal and expected real yields are likely to be very highly correlated, which suggests that his estimates are appropriate for an IS curve that depends on the expected real rate.

[8] To make the argument in the preceding two paragraphs concrete, it may help to calculate y_t's ARMA parameters for specific λ and a. Suppose that $a = .96$, $\lambda = .96$. (The value for λ is Fama and Gibbons's (1982, Table 2) point estimate of the first-order serial correlation coefficient of monthly T-bill rates, 1953–77; Fama and Gibbons do not report figures for annual interest rates.) Then one can grind through the formulas in the text to show that y_t is moving average polynomial factors as $(1+.32L)(1-.98L)$. Output will therefore behave much like an ARMA(1,1) variable with a single-autoregressive unit root of .96.

More generally, for any λ between 0 and 1, there will also be persistence in GNP. If λ is near zero, y_t will behave much like the serially uncorrelated variable defined in (9). The closer λ is to unity, the more y_t will behave like the serially correlated process defined in (10).

It is worth noting that a similar result obtains if, as in Taylor (1980a, b) the money supply rule involves targeting the money supply instead of the interest rate. To analyze this type of rule, it is convenient to follow Taylor and replace the IS and LM curves with a simple quantity equation,

$$y_t + p_t = m_t. \tag{11}$$

Unlike Taylor, I have set to zero shocks to velocity (deviations of velocity from trend), as explained in the introduction. Also, replace the interest rate target (5) with a money supply target,

$$m_t = gp_t + \lambda(m_{t-1} - gp_{t-1}) + u_t, \tag{12}$$

where u_t is a serially uncorrelated shock and $0 \leq \lambda \leq 1$.

To understand (12), consider first the case when $\lambda, u_t \equiv 0$. Then, as in Taylor (1980a), the parameter g measures how accommodative monetary policy is. (For my purposes one could have the monetary authority look directly at y_t as well as, or instead of, p_t; only p_t appears in the money supply rule for consistency with Taylor (1980a, b).) The shock u_t is not present in Taylor (1980a, b). It is intended to reflect shocks to the money supply resulting from, say, random movements in the money multiplier. The $\lambda(m_{t-1} - gp_{t-1})$ term is present to capture a tendency of the monetary authority to absorb previous control errors. If $\lambda = 1$, previous u_t's are never offset and are carried through to all future money supplies. Such random-walk behavior ("base drift") has been argued to characterize Federal Reserve policy in the United States, at least in recent years (see Carl Walsh, 1986, and the references cited therein).

The model may be solved as in Taylor (1980a, b); the details are omitted to save space. Let $\beta = 1 - g$; $c = (1 + .5\beta\gamma)(1 - .5\beta\gamma)^{-1}$; $a = c - (c^2 - 1)^{1/2}$, if $c > 1$; $a = c + (c^2 - 1)^{1/2}$, if $c < -1$; $b = .5\gamma(\lambda + \lambda^2)/[1 + .5\beta\gamma - .5(1 - .5\beta\gamma)(a + \lambda)]$. Then

$$
\begin{aligned}
(13)\quad x_t &= ax_{t-1} + b(m_{t-1} - gp_{t-1}),\\
y_t &= -.5\beta b[(1+L)/(1-aL)]\\
&\quad\times(m_{t-1} - gp_{t-1})\\
&\quad+\lambda(m_{t-1} - gp_{t-1}) + u_t.
\end{aligned}
$$

Consider first the case where $\lambda = 0$, $m_t = gp_t + u_t$. Since $\lambda = 0$ implies $b = 0$ (see the formula for b above equation (13)), we have $y_t = u_t$, and y_t is serially uncorrelated. Suppose instead that $\lambda = 1$, and $m_t = gp_t + (m_{t-1} - gp_{t-1}) + u_t$. Then it is straightforward but tedious to show that (13) reduces to

$$y_t = ay_{t-1} + u_t + .5(1-a)u_{t-1}.$$

So $y_t \sim$ ARMA (1,1). In any finite sample, y_t will look arbitrarily like a random walk for $\underline{a}$ arbitrarily close to unity. Now, $a \rightarrow 1$ as $\beta\gamma = (1-g)\gamma \rightarrow 0$. As was just noted, γ is about .01 to .10. Taylor (1980b) estimated β to be about .3. This indicates that $\beta\gamma$ is about .003 to .03, yielding an implied $\underline{a}$ of .78 to .93. See Table 1, Part B.

The intuition to the effect of λ on the univariate y_t process is similar to that for the previous money supply rule. With $\lambda = 0$, the contract wage and price level are nonstochastic: $x_t = x_{t-1} = 0$ is the only stationary solution to (13) with $b = 0$. So the aggregate demand equation (11) implies $y_t = m_t$, and with $\lambda = 0$, m_t is serially uncorrelated. By contrast, when $\lambda = 1$, the unit root in the money supply first of all puts a unit root in wages and prices; the long-run properties of wages and prices are governed by the money supply. But because of the staggered contracts, real balances, the difference between m and p, have additional persistence: prices do not move instantaneously, and one for one, with money supplies. This persistence is transmitted directly into GNP by the aggregate demand equation (11).

II. Conclusions

Neither stationarity of the natural rate nor nominal shocks playing an important role in the business cycle are inconsistent with a root very near to unity being present in the GNP process. In Taylor's (1980a, b) stationary natural rate model, extreme persistence in GNP movements is precisely what is predicted, given stylized versions of money supply rules often attributed to the Fed, and plausible values for the model's basic parameters. The model also predicts that different money supply rules would result in dramatically less persistent movements in GNP.

This is not to argue that, in fact, the business cycle in the United States is purely, or even largely, monetary in origin, or that natural rate theory is to be preferred to nonnatural rate theory. Rather, detailed study of the univariate process for a single country's GNP is unlikely to be particularly helpful in deciding some important business cycle issues. Potentially more helpful are comparative studies of GNP processes across various countries and various time periods. The evidence here is mixed. Stultz and Wasserfallen (1985) and Campbell and Mankiw (1987) conclude that during the postwar period the random-walk approximation is reasonable for a number of industrialized countries. This perhaps makes it less likely that GNP behavior could change dramatically with a change in policy regime. On the other hand, Stock and Watson (1986) find that for the United States, the random-walk approximation is reasonable only in the post-1919 period. This is consistent with the present paper's model: the near random-walk behavior of the nominal interest rate appears to have begun around 1915–1920 (Mankiw et al., 1987), and inflation appears to have been more sensitive to excess demand pre-1929 than postwar (Sachs, 1980).[9] In any case, estimation of multivariate structural models is, of course, potentially still more helpful than estimation of univariate time-series.

APPENDIX

As stated in fn. 4, the money supply rule (5) is understood to be the limit of a certain non-interest rate rule. This rule is a simple generalization of the rule in Robert Driskell and Steven Sheffrin, 1986, and Bradford De Long and Lawrence Summers, 1986:

$$m_t = \alpha(i_t - z_t), \quad z_t \equiv u_t/(1-\lambda L), \quad \alpha > 0. \tag{A1}$$

Thus, if i_t is above (below) its target level z_t, m_t is increased (decreased). The rule (A1) yields a unique stationary solution for y_t for any finite α; the solution for y_t under the rule (5) is understood to be the one that results when one first solves using a finite α and then takes the limit as $\alpha \rightarrow \infty$.

Use (A1) to eliminate m_t from the LM curve (4) and rearrange to get

$$i_t = (\alpha + \theta_2)^{-1}(p_t + \alpha z_t + \theta_1 y_t).$$

Substitute the above into the IS curve (3) and rearrange to get

$$\begin{aligned} y_t = {} & \left[1 + (\alpha+\theta_2)^{-1}\delta\theta_1\right]^{-1} \\ & \times\Big\{-\delta\alpha(\alpha+\theta_2)^{-1}z_t \\ & \quad + \delta_t p_{t+1} - \delta\left[1+(\alpha+\theta_2)^{-1}\right]p_t\Big\} \\ \equiv {} & -\delta_0 z_t + \delta_{1t} p_{t+1} - \delta_2 p_t. \end{aligned} \tag{A2}$$

---

[9] Stock and Watson (1986) suggest that the seeming stationarity of pre-1919 GNP may instead be an artifact of the way these data were constructed.

Since ${}_{t-1}z_t = \lambda z_{t-1}, {}_{t-1}z_{t+1} = \lambda^2 z_{t-1}$, (A2) implies

$$\text{(A3)}\quad {}_{t-1}y_t + {}_{t-1}y_{t+1} = -\delta_0(\lambda+\lambda^2)z_{t-1} + \delta_1({}_{t-1}p_{t+2} + {}_{t-1}p_{t+1}) - \delta_2({}_{t-1}p_{t+1} + {}_{t-1}p_t).$$

Use the price markup equation (2) to eliminate the price terms from (A3) and substitute the result into the supply equation (1). After some rearrangement, this becomes

$$\text{(A4)}\quad (\gamma\delta_1)\,{}_{t-1}x_{t+2} + [2+\gamma(2\delta_1-\delta_2)]\,{}_{t-1}x_{t+1} - [4+\gamma(2\delta_2-\delta_1)]\,{}_{t-1}x_t + (2-\gamma\delta_2)x_{t-1} - 4(x_t - {}_{t-1}x_t) = 2\gamma(\lambda+\lambda^2)\delta_0 z_{t-1}.$$

For (A4) to hold, $x_t - {}_{t-1}x_t$ must be identically zero. It follows from Driskell and Sheffrin (1986) that the polynomial

$$\text{(A5)}\quad (\gamma\delta_1)a^3 + [2+\gamma(2\delta_1-\delta_2)]a^2 - [4+\gamma(2\delta_2-\delta_1)]a + (2-\gamma\delta_2)$$

has exactly one stable and two unstable roots. Let the unique stable root be a_1. Since $z_t \sim$ AR(1), it follows that solving the unstable roots forward, the stable root backward leads to a solution of the form $x_t = a_1 x_{t-1} + b_1 z_{t-1}$. One can solve for b_1 by using $x_t = a_1 x_{t-1} + b_1 z_{t-1}$ to compute ${}_{t-1}x_{t+2}, {}_{t-1}x_{t+1}$ and ${}_{t-1}x_t$, and putting these into (A4). (The exact formula is not of interest.)

Let $\alpha \to \infty$. (The solution for y_t is the same whether one uses the present technique of solving for y_t using the x_t process that results for $\alpha \to \infty$, or solves for y_t for finite α and then lets $\alpha \to \infty$.) Then $\delta_0, \delta_1, \delta_2 \to \delta$ and (A5) reduces to

$$(a-1)[\gamma\delta a^2 + (2+2\gamma\delta)a - (2-\gamma\delta)].$$

Since $|a_1| < 1$ for finite α, then, by continuity, as $\alpha \to \infty$, a_1 approaches the stable root of $\gamma\delta a^2 + (2+2\gamma\delta)a - (2-\gamma\delta)$. This is $\underline{a} \equiv -(1+\gamma^{-1}\delta^{-1}) + \gamma^{-1}\delta^{-1}(4\gamma\delta+1)^{1/2}$. Also $b_1 \to b$, where b is given in equation (6). Equation (7) now follows.

It is perhaps worth noting that one can derive the same result concerning near random-walk behavior of y_t by letting $\lambda = 1$, but (a) assuming that the α in equation (A1) is finite (but large), or (b) letting $\alpha \to -\infty$ rather than $\alpha \to +\infty$.

REFERENCES

Blanchard, Olivier J. and Quah, Danny, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," unpublished manuscript, MIT, 1987.

Campbell, John Y. and Mankiw, N. Gregory, "Are Output Fluctuations Transitory?," unpublished manuscript, 1986.

______ and ______, "International Evidence on the Persistence of Economic Fluctuations," unpublished manuscript, Harvard University and Princeton University, 1987.

Clark, Peter K., (1986a) "The Cyclical Component of U.S. Economic Activity," unpublished manuscript, Stanford University, 1986.

______, (1986b) "The Shape of U.S. Business Cycles," unpublished manuscript, Stanford University, 1986.

Cochrane, John H., "How Big Is the Random Walk in GNP?," unpublished manuscript, University of Chicago, 1986.

Deaton, Angus, "Life Cycle Models of Consumption: Is the Evidence Consistent with the Theory?," NBER Working Paper No. 1910, 1986.

De Long, J. Bradford and Summers, Lawrence H., "Is Increased Price Flexibility Destabilizing?," *American Economic Review*, December 1986, *76*, 1031–44.

Dickey, David A. and Fuller, Wayne A., "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econo-*

metrica, June 1981, *49*, 1057–72.

Driskell, Robert A. and Sheffrin, Steven M., "Is Price Flexibility Destabilizing?," *American Economic Review*, September 1986, *76*, 802–07.

Fama, Eugene F. and Gibbons, Michael, "Inflation, Real Returns and Capital Investment," *Journal of Monetary Economics*, May 1982, *9*, 297–323.

Friedman, Benjamin M., "The Inefficiency of Short-Run Monetary Targets for Monetary Policy," *Brookings Papers on Economic Activity*, 1:1977, 293–335.

Fuller, Wayne A., *Introduction to Statistical Time Series*, New York: Wiley & Sons, 1976.

King, Robert, Plosser, Charles, Stock, James and Watson, Mark, "Stochastic Trends and Economic Fluctuations," NBER Working Paper No. 2229, 1987.

Mankiw, N. Gregory, "The Optimal Collection of Seignorage: Theory and Evidence," unpublished manuscript, Harvard University, 1986.

______, **Miron, Jeffrey A. and Weil, David N.**, "The Adjustment of Expectations to a Change in Regime: A Study of the Founding of the Federal Reserve," *American Economic Review*, June 1987, *77*, 358–74.

McCallum, Bennett T., "A Note on John Taylor's Macroeconomic Model," unpublished manuscript, Carnegie-Mellon University, undated.

______, (1986a) "On 'Real' and 'Sticky Price' Theories of the Business Cycle," *Journal of Money, Credit, and Banking*, November 1986, *28*, 397–414.

______, (1986b) "Some Issues Concerning Interest Rate Pegging, Price Level Determinacy, and the Real Bills Doctrine," *Journal of Monetary Economics*, January 1986, *17*, 135–60.

______, "Inflation: Theory and Evidence," unpublished manuscript, Carnegie-Mellon University, 1987

Nelson, Charles R. and Plosser, Charles I., "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, September 1982, *10*, 139–62.

Quah, Danny, "What Do We Learn from Unit Roots in Macroeconomic Time Series?," unpublished manuscript, MIT, 1986.

Rose, Andrew K., "The Autoregressivity Paradox," unpublished manuscript, University of California-Berkeley, 1987.

Sachs, Jeffrey, "The Changing Cyclical Behavior of Wages and Prices: 1890–1976," *American Economic Review*, March 1980, *70*, 78–90.

Sargent, Thomas J., *Macroeconomic Theory*, New York: Academic Press, 1979.

Stock, James H. and Watson, Mark W., "Does GNP have a Unit Root?," *Economic Letters* 1986, *22*, 147–51.

Stultz, Rene M. and Wasserfallen, Walter, "Macroeconomic Time Series, Business Cycles, and Macroeconomic Policies," in K. Brunner and A. Meltzer, eds., *Understanding Monetary Regimes*, Vol. 22, *Carnegie-Rochester Conference Series on Public Policy*, *J. Monet. Econ.*, Suppl. 1985, 9–55.

Taylor, John B., (1980a) "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, February 1980, *88*, 1–23.

______, (1980b) "Output and Price Stability: An International Comparison," *Journal of Economic Dynamics and Control*, February 1980, *2*, 109–32.

______, "What Would Nominal GNP Targeting Do to the Business Cycle?," in K. Brunner and A. Meltzer, eds., *Understanding Monetary Regimes*, Carnegie-Rochester Conference Series on Public Policy, *22*, Spring 1985, *J. Monet. Econ.* Suppl., 61–84.

Walsh, Carl, "In Defense of Base Drift," *American Economic Review*, September 1986, *76*, 692–700.

Watson, Mark W., "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, July 1986, *18*, 49–76.

Determinants of the Equilibrium Real Exchange Rate

By PETER NEARY*

The importance of the equilibrium real exchange rate, the relative price of nontraded to traded goods consistent with balance-of-payments equilibrium, has long been recognized.[1] However, no simple, compact derivation of the determinants of the real exchange rate in a multicommodity framework appears to be available.[2] Section I below attempts to fill this gap, and Section II notes the implications of the main result for a number of applied questions.

I. Theory

I assume a small open economy producing arbitrary numbers of traded and nontraded goods under competitive conditions. The world prices of traded goods are exogenously given, and, without loss of generality, I take one of the traded goods as numeraire so that domestic prices are measured in terms of foreign currency. Since I consider only situations of balance-of-payments equilibrium, it is immaterial whether this is achieved by adjustment of a floating exchange rate or by monetary inflows and outflows under a fixed exchange rate. I assume that there are n nontraded goods, whose prices, denoted by the vector p, are endogenously determined so as to equate domestic demand and supply. The determination of equilibrium may be illustrated in a compact manner by making use of the *trade expenditure function*.[3] This equals the excess of home expenditure, determined by a standard expenditure function, $e(p, u, \alpha)$, over home income from production, determined by a GNP function, $g(p, \beta)$, plus any additional income received by the private sector, γ. In symbols:

$$(1) \qquad E(p,u,\phi)=e(p,u,\alpha)-g(p,\beta)-\gamma.$$

Here $\phi=(\alpha, \beta, \gamma)$ is a vector of shift parameters whose interpretation will vary with the particular shocks under consideration. For example, α may include the prices of nonnumeraire traded goods as well as taste parameters; β may include both factor endowments and technology parameters; and γ may include redistributed tax or tariff revenue as well as direct transfers from abroad. The scalar u measures home utility, which can be interpreted either as the actual utility of a single aggregate consumer or as the level of social welfare attained when optimal redistribution is continually carried out by means of lump-sum transfers. Relaxing this assumption to allow for different individuals in the economy with optimal lump-sum redistribution not carried out would introduce additional complications that, though potentially important in practice, are not central to the applications of the model discussed below. In any case, these considerations have been widely discussed in other contexts.[4]

---

*University College, Belfield, Dublin 4, Ireland; and Queen's University, Kingston, Ontario K7L 3N6 Canada. An earlier version of this paper was circulated as Discussion Paper No. 209 of the Centre for Economic Policy Research, London, and was presented at the Econometric Society European Meetings in Copenhagen in August 1987. I am grateful for helpful comments to participants on that occasion and to Mick Devereux, Lars Svensson, Scott Taylor, Stephen Yeo, and a referee.

[1] Peter Oppenheimer (1974) documents early writings on models with nontraded goods, whereas W. E. G. Salter (1959) and Michael Bruno (1976) are representative of the voluminous modern literature.

[2] By contrast, the building blocks of such a derivation are well known. See, in particular, Richard Caves and Ronald Jones (1985, pp. 494–98), Avinash Dixit and Victor Norman (1980, ch. 5), and Alan Woodland (1982, chs. 8 and 10).

[3] The detailed properties of this function have been examined by Dixit and Norman (1980, pp. 90–91), who call it the "excess expenditure function," and Woodland (1982, p. 170), who calls its negative the "maximum net revenue function."

[4] To give only two examples, Dixit and Norman (1986) examine the problems that arise in measuring

The great convenience of the trade expenditure function is that, for fixed ϕ, its properties resemble those of the usual expenditure function, $e(p,u,\alpha)$. In particular, it is increasing in u, it is concave in p, and its partial derivatives with respect to p equal the Hicksian or utility-compensated excess demand functions for nontraded goods:[5]

$$(2)\quad E_p(p,u,\phi)=e_p(p,u,\alpha)-g_p(p,\beta).$$

The function E is homogeneous of degree zero in the prices of all goods, traded as well as nontraded. However, since p denotes the prices of nontraded goods only, the matrix of second partial derivatives of E with respect to p will be nonsingular provided there is some substitutability in either demand or supply between traded and nontraded goods. I assume this henceforward and denote the negative of this matrix by S:

$$(3)\qquad S=-E_{pp}(p,u,\phi).$$

Because of the properties of e and g, S is symmetric and positive definite.

It is now straightforward to characterize equilibrium. First, balance of payments equilibrium requires that aggregate expenditure equal aggregate income:[6]

$$(4)\qquad E(p,u,\phi)=0.$$

Second, commodity market equilibrium requires that excess demand for each nontraded good be zero:

$$(5)\qquad E_p(p,u,\phi)=0.$$

The $n+1$ equations (4) and (5) determine the values of the $n+1$ endogenous variables, p and u, as functions of the parameters ϕ.

To solve the model, first totally differentiate (4) to obtain

$$(6)\qquad du=-E'_\phi\,d\phi\equiv d\Phi.$$

Here, I have used (5) to eliminate a term in dp and have chosen units of measurement for utility such that $E_u=e_u=1$. I have also introduced the scalar parameter Φ to denote the aggregate effect of any given combination of exogenous shocks: $d\Phi$ measures the net change in utility arising from changes in all the exogenous variables. Next, totally differentiate (5) to obtain:

$$(7)\qquad \begin{aligned} -S\,dp+E_{pu}\,du&=-E_{p\phi}\,d\phi\\ &\equiv -E_{p\Phi}\,d\Phi.\end{aligned}$$

The two terms on the left-hand side of (7) measure respectively the substitution effects of price changes and the income effects of utility changes on the excess demand for nontraded goods. The right-hand side measures the effects of changes in exogenous variables on excess demand; as in (6), it is convenient to aggregate these effects, with each element of the vector $E_{p\Phi}$ giving the net increase in excess demand for each nontraded good as a result of changes in exogenous variables.[7] Combining (6) and (7) and solving for the vector of changes in nontraded goods prices yields

$$(8)\qquad dp=S^{-1}(E_{pu}+E_{p\Phi})\,d\Phi.$$

The final step is to aggregate the price changes of individual nontraded goods to obtain the change in the real exchange rate. I define the latter as a fixed-weight index number of the relative prices of nontraded goods divided by a price index for traded goods,

---

welfare changes when lump-sum redistribution is not feasible, and Lance Taylor (1974) considers the positive implications of intranational differences in spending propensities in a model with traded and nontraded goods.

[5] Throughout, terms such as E_p denote the vector of partial derivatives of E with respect to p, and terms such as E_{pp} denote the matrix of second partial derivatives. All vectors are column vectors, and a prime denotes a transpose.

[6] In a multiperiod model, a payments imbalance could be consistent with full intertemporal equilibrium. I ignore such considerations in what follows.

[7] Thus, the ith component of the vector $E_{p\Phi}$ equals $-\sum_j E_{ij}\,d\phi_j/\sum_j E_j\,d\phi_j$, where E_{ij} is shorthand for $\partial^2E/\partial p_i\partial\phi_j$.

p_T:[8]

$$(9) \qquad \pi \equiv x'p/p_T,$$

where the elements of the vector x are the base-period consumption (and production) levels of individual nontraded goods. Totally differentiating, with p_T assumed constant, and combining with (8) gives the principal result of this note:

$$(10) \quad d\pi = x'S^{-1}(E_{pu} + E_{p\Phi})\, d\Phi/p_T.$$

The interpretation of this result is facilitated if the parameters are expressed in proportional terms. This may be done by introducing a diagonal matrix P, the ith element on the principal diagonal of which is the price of the ith nontraded good. Straightforward manipulations now allow (10) to be rewritten as follows:

$$(11) \qquad \hat{\pi} = \omega'\bar{S}^{-1}(\mu - \sigma)\, d\Phi.$$

Here $\hat{\pi}$ equals $d\pi/\pi$; ω equals $Px/x'p$, the vector of individual commodity weights in the nontraded goods price index; $\bar{S}$ equals PSP, a simple transformation of the substitution matrix defined in (3); μ equals PE_{pu}, the vector of *marginal propensities to consume* nontraded goods; and σ equals $-PE_{p\Phi}$, which may be interpreted as the vector of *marginal propensities to produce* nontraded goods. Of course, the values of the latter depend on the particular shock under consideration, as the examples in the next section illustrate.

II. Applications

Equation (11) may be interpreted as implying that any exogenous shock will raise π and so lead to a *real appreciation* if the demand effects of the shock outweigh the supply effects. The coefficient of $d\Phi$ in (11) is a weighted sum of the differences between the marginal propensities to consume and to produce each nontraded good. The weights equal the elements of the vector $\omega'\bar{S}^{-1}$ and so reflect both the relative importance of each good in consumption and the difficulty of substituting it for other goods. While there is no guarantee that all these weights must be positive,[9] the fact that $\bar{S}$ is a positive definite matrix justifies the statement that a real appreciation will occur provided the elements of ω and $(\mu - \sigma)$ are positively correlated. The significance of this result is best seen by considering a number of applications.

A. *Effects of a Transfer*

An incoming transfer, such as foreign aid, is necessarily effected in terms of traded goods. Hence, at initial prices, it does not affect the output of nontraded goods and so all the elements of σ are zero. It follows (assuming nontraded goods are normal in demand) that a transfer is likely to induce a real appreciation, a prediction that has been confirmed empirically by Michael Michaely (1981).

B. *The "Dutch Disease"*

A sector-specific boom has effects identical to those of a transfer if it occurs in an "enclave" traded-good sector, which has no production links with the rest of the economy. (Natural resource sectors are an obvious example.) Such a boom has a "spending effect" only (in the terminology of Max Corden and myself [1982]) and necessarily leads to a real appreciation. Real national income and so potential welfare rises, although the loss of competitiveness experienced by nonbooming traded sectors may pose problems of adjustment and may lead to loss of learning-by-doing gains in those sectors. These potentially harmful side effects of a boom have come to be known as the "Dutch Disease." The effect on the real exchange rate is less clear-cut if the booming sector is integrated with the rest of the economy so that the boom also has a "resource movement effect." In particular, if the boom, through general-equilibrium interactions, encourages more production of nontraded goods at initial prices (i.e., the elements of $g_{p\beta}$ and hence of σ are positive), then a real appreciation is not inevitable.[10] Expressed in the terminology of Sir John Hicks (1953), the more home-market biased is the sectoral pattern of economic growth, the less likely is a real appreciation. In the neutral benchmark case of "balanced growth" and homothetic tastes, μ_i and σ_i are equal for each nontraded good; hence all relative prices, including the real exchange rate, are unaffected by growth.

---

[8] These price indices may be given a rigorous justification for large as well as small changes if the trade expenditure function is homothetically separable in the prices of traded and nontraded goods.

[9] One special case in which the weights must be positive is where all goods are substitutes in aggregate excess demand. Then S has positive diagonal entries and negative off-diagonal entries, and (from the homogeneity restrictions on excess demand) SP has a dominant diagonal. It follows that all the elements of $(SP)^{-1}$ are positive, which implies that the same must be true of the elements of $\bar{S}^{-1}$.

C. *International Comparisons of Purchasing Power Parity*

The doctrine of purchasing power parity may be expressed as the prediction that price levels should be equal across countries when compared by means of equilibrium exchange rates. However, there is ample evidence against this hypothesis and in favor of the view that nontraded goods are relatively cheaper in low-income countries. (See Irving Kravis, Alan Heston, and Robert Summers, 1978.) Bela Balassa (1964) and Paul Samuelson (1964) argue that such deviations from purchasing power parity are to be expected because higher-income countries have higher relative productivity in the production of traded goods. This implication follows clearly from (11): $\hat{\pi}$ and $d\Phi$ may be reinterpreted as the differences between a high-income and a low-income country in the real exchange rate and the level of real income respectively, and μ and σ may be reinterpreted as the differences between the two countries in *average* propensities to consume and produce nontraded goods at constant prices. If international productivity differences are smaller in the production of nontraded goods (such as services) than in the production of traded goods, the elements of σ are likely to be small or even negative,[11] and so the higher-income country will have a higher real exchange rate (a higher relative price of services). This effect will be reinforced if nontraded goods are superior in demand (so that the elements of μ are positive and relatively large). Although cross-section evidence in favor of this productivity-differential hypothesis is not strong (see Balassa, 1964, and Lawrence Officer, 1976), David Hsieh (1982) shows that it is confirmed by time-series tests.[12] His suggestion that this occurs because country-specific factors such as tastes vary relatively little over time is fully in keeping with equation (11).

D. *Changes in the Terms of Trade*

Although I have so far assumed that the prices of all traded goods are given, $d\Phi$ may be interpreted as the change in income arising from a change in the terms of trade. Suppose that export prices rise so that the terms of trade improve. In this application μ denotes the marginal propensities to consume nontraded goods as before, but σ must be reinterpreted as the cross-price effects on excess supply of nontraded goods of the terms of trade improvement. These must be negative if every traded good is a substitute for every nontraded good, thus reinforcing the income effect. This gives the conclusion that, except where complementary relationships dominate, a terms of trade improvement tends to lead to a real appreciation.[13]

---

[10] See Peter Neary and Sweder van Wijnbergen (1986) for further theoretical and empirical elaboration.

[11] For example, in a two-sector model, if the higher-income country has a Hicks-neutral technological superiority in the traded sector only, then it must produce relatively less of the nontraded good at given prices (i.e., σ must be negative) in either the Heckscher-Ohlin or specific-factors model. With more general forms of technological differences, σ may be positive (see Ronald Findlay and Harry Grubert, 1959; Ronald Jones, 1970; and Neary, 1981) but there is still a presumption that it will be low or negative.

[12] Of course, the results of such tests should be interpreted with care, since strictly speaking the hypothesis applies only to steady-state comparisons in which balance of payments equilibrium holds continually.

III. Conclusion

The approach adopted in this paper shows that a number of results in the literature on small open economies are special cases of a general phenomenon: changes in exogenous variables are more likely to lead to a real appreciation the greater their effect on the demand for and the smaller their effect on the supply of nontraded relative to traded goods. Unfortunately, this paper throws no light on one of the most difficult problems facing all students of the open economy—theorists, applied economists, and policymakers alike—namely, determining how far actual real exchange rates diverge from their equilibrium values.

---

[13] Rudiger Dornbusch (1980, ch. 6) gives a three-commodity exposition of this result. If the terms of trade improvement arises from a fall in import prices, the income and substitution effects work in opposite directions. Since σ_i equals $p_i \Sigma_j E_{ij}\, dq_j / \Sigma_j E_j\, dq_j$, where the denominator (which equals $-d\Phi$) is negative following a terms of trade improvement and the E_{ij} terms are positive by assumption, the sign of σ_i depends on the sign of the price changes dq_j.

REFERENCES

Balassa, Bela, "The Purchasing-Power Parity Doctrine: A Reappraisal," *Journal of Political Economy*, December 1964, *72*, 584–96.

Bruno, Michael, "The Two-Sector Open Economy and the Real Exchange Rate," *American Economic Review*, September 1976, *66*, 566–77.

Caves, Richard E. and Jones, Ronald W., *World Trade and Payments*, 4th ed., Boston: Little, Brown, 1985.

Corden, W. Max and Neary, J. Peter, "Booming Sector and Deindustrialisation in a Small Open Economy," *Economic Journal*, December 1982, *92*, 825–48.

Dixit, Avinash K. and Norman, Victor, *Theory of International Trade: A Dual, General Equilibrium Approach*, Welwyn: James Nesbit, 1980.

______ **and** ______, "Gains from Trade without Lump-Sum Compensation," *Journal of International Economics*, August 1986, *21*, 111–22.

Dornbusch, Rudiger, *Open-Economy Macroeconomics*, New York: Basic Books, 1980.

Findlay, Ronald and Grubert, Harry, "Factor Intensities, Technological Progress, and the Terms of Trade," *Oxford Economic Papers*, February 1959, *11*, 111–21.

Hicks, John R., "An Inaugural Lecture: 2. The Long-Run Dollar Problem," *Oxford Economic Papers*, June 1953, *5*, 121–35.

Hsieh, David A., "The Determination of the Real Exchange Rate: The Productivity Approach," *Journal of International Economics*, May 1982, *12*, 355–62.

Jones, Ronald W., "The Role of Technology in the Theory of International Trade," in R. Vernon, ed., *The Technology Factor in International Trade*, New York: NBER, 1970, 73–92.

Kravis, Irving, Heston, Alan and Summers, Robert, *International Comparisons of Real Product and Purchasing Power*, Baltimore: Johns Hopkins Press, 1978.

Michaely, Michael, "Foreign Aid, Economic Structure and Dependence," *Journal of Development Economics*, December 1981, *9*, 313–30.

Neary, J. Peter, "On the Short-Run Effects of Technological Progress," *Oxford Economic Papers*, July 1981, *33*, 224–33.

______ **and van Wijnbergen, Sweder,** eds., *Natural Resources and the Macroeconomy*, Oxford: Basil Blackwell and Cambridge, MA: MIT Press, 1986.

Officer, Lawrence H., "The Productivity Bias in Purchasing Power Parity: An Econometric Investigation," *International Monetary Fund Staff Papers*, November 1976, *23*, 545–79.

Oppenheimer, Peter, "Non-Traded Goods and the Balance of Payments: A Historical Note," *Journal of Economic Literature*, September 1974, *12*, 882–88.

Salter, W. E. G., "Internal and External Balance: The Role of Price and Expenditure Effects," *Economic Record*, August 1959, *35*, 226–38.

Samuelson, Paul A., "Theoretical Notes on Trade Problems," *Review of Economics and Statistics*, May 1964, *46*, 145–54.

Taylor, Lance, "Short-Term Policy in Open Semi-Industrialized Countries: The Narrow Limits of the Possible," *Journal of Development Economics*, September 1974, *1*, 85–104.

Woodland, Alan D., *International Trade and Resource Allocation*, Amsterdam: North-Holland, 1982.

Nonadditive Preferences and the Marginal Propensity to Consume

By EHUD I. RONN*

Many authors have made the simplifying assumption of a time-additive utility function. Examples include Edmund Phelps, 1962; James Mirrlees, 1965; Nils Hakansson, 1970; and Robert Merton, 1971. The advantage of time additivity follows from its analytical tractability. Time additivity, however, incorporates the implicit assumption of zero complementarity across time. It is thus of interest to see what effects the relaxation of their assumption yields.[1]

Some of the earliest work concerning the extension of consumption-investment decisions from certainty to uncertainty was carried out by Phelps (1962) and Mirrlees (1965). Both authors considered isoelastic multiperiod time-additive utility functions with stationary investment opportunity sets. Their results were extended by D. Levhari and T. N. Srinivasan (1969) and Leonard Mirman (1971); the former incorporated a stationary investment opportunity set to derive the optimal consumption-investment decision while the latter considered a two-period model.

Paul Samuelson (1969) analyzed discrete-times, time-additive logarithmic and isoelastic utility functions. His results were extended and generalized in discrete time by Hakansson (1970) and in continuous time by Merton (1971). Further, Alan Kraus and Robert Litzenberger (1975) and Mark Rubinstein (1976) derived market equilibrium implications of identical economy-wide logarithmic preferences. Their analyses, however, fail to account for any time complementarity (in that their posited utility functions are time additive), and that is one of the objectives of this paper.

Hayne Leland (1968) considered the effect on saving of the *uncertainty* of future income, the "precautionary" demand for saving, in the context of a two-period model with arbitrary preferences. In contrast, the current paper investigates the effect of larger current *certain* wealth on the marginal propensity to consume. Leland found that risk aversion per se was not a sufficient condition for greater savings in the presence of increased uncertainty. Rather, additional assumptions were required, such as time-additive preferences displaying decreasing absolute risk aversion. The current paper demonstrates that, under nonadditive preferences, risk aversion is not a sufficient condition for a *positive* marginal propensity to consume.

Scott Richard (1975) defined multivariate risk aversion over two or more attributes:

> Let $u(x, y)$ be a decision-maker's utility function for two attributes [where $u_x > 0, u_y > 0$]. If for any $x_0 < x_1$ and any $y_0 < y_1$, the decision maker prefers lottery two which gives an even chance for (x_0, y_1) or (x_1, y_0) to lottery one which gives an even chance for (x_0, y_0) or (x_1, y_1), then the decision maker is considered multivariate risk averse [p. 12].

Richard then demonstrated that a necessary and sufficient condition for (strict) multivariate risk aversion is the negativity of the second cross-partial derivative of the utility function with respect to any two arguments of the utility function. Richard did not, how-

---

*Schools of Business Administration, University of California-Berkeley, CA 94720, and University of Chicago, Graduate School of Business, Chicago, IL 60637. The author gratefully acknowledges the helpful suggestions of Nils Hakansson, Hayne Leland, Scott Richard, Mark Rubinstein, and an anonymous referee, while remaining solely responsible for any errors contained herein.

[1] Thus, for the purposes of this paper, (positive/negative) time complementarity is definitionally equivalent to nonadditivity in intertemporal consumption choices.

ever, analyze the consumptive implications of this form of risk aversion, and that is the objective of this paper.

The main result of this paper is as follows. Consider the marginal propensity to consume of an individual at his/her consumptive-investment optimum. If this individual is sufficiently multivariate risk averse (in the Richard sense), and if the investment opportunity set is sufficiently disadvantageous, then the individual's marginal propensity to consume (MPC) will be negative. If the individual is a multivariate risk preferrer ($\partial^2 U/\partial C_t \partial C_\tau > 0$ for $t \neq \tau$) or risk neutral (displays time-additive preferences), then the MPC is positive and its positivity is independent of investment opportunities. Thus, a startling result obtains: the intuitively appealing feature of multivariate risk aversion can, in the presence of "low" (real) interest rates, give rise to negative MPC's.

Section I below presents the intuition underlying multivariate risk aversion while Section II derives the result cited above. Concluding thoughts are contained in Section III.

I. Multivariate Risk Aversion and Nonadditive Preferences

This section analyzes the relationship of nonadditive preferences to the marginal propensity to consume in the context of complete (Arrow-Debreu) capital markets. One of the appealing features of the current result is that it provides a natural definition and role for a multivariate extension of the well-known coefficients of absolute and relative risk aversion. Thus, define the coefficient of multivariate risk aversion, MVRA, as

$$(1)\qquad \text{MVRA}(s) \equiv -\frac{\partial^2 U}{\partial C_0 \partial C_s} \bigg/ \frac{\partial U}{\partial C_s}$$

for a given state of the world s. This definition of MVRA(s) as given in equation (1) may be justified by noting that sign MVRA $= -\text{sign}\,(\partial^2 U/\partial C_0 \partial C_s)$ since nonsatiety assures that $\partial U/\partial C_s > 0$. Thus, MVRA > 0 implies multivariate aversion to risk and MVRA $\equiv 0$ is synonymous with time-additive utilities. Further, the well-known coefficient of absolute risk aversion, $-(\partial^2 U/\partial C_s^2)/(\partial U/\partial C_s)$, similarly features the (own) second derivative divided by marginal utility, and both coefficients are invariant to increasing linear transformations of the utility function.

Moreover, in analogy to the coefficient of absolute risk aversion, the MVRA as defined in equation (1) can be shown to prescribe the magnitude of the risk premium an individual would pay to avoid (an infinitesimally small) gamble. Thus, proceeding in a multivariate extension of John Pratt's (1964) seminal analysis, define

$$(2)\qquad U(C_0^*, C_1^*, \dots, C_n^*) \equiv \max_{\{C_0, C_1, \dots, C_n\}} E\left[U(C_0, \tilde{C}_1)\right],$$

subject to a budget constraint, where C_0 is current consumption and C_i is state-contingent consumption in state i (of period 1). Now fix $(C_2^*, \dots, C_n^*)$, so that $v(C_0^*, C_1^*) \equiv U(C_0^*, C_1^*, \dots, C_n^*)$, and define the state 1 risk premium p as the implicit solution to

$$(3)\qquad v(C_0^*, C_1^* - p) = E\left[v(C_0^* + \tilde{z}, C_1^* + \tilde{z})\right],$$

where $E(\tilde{z}) = 0 < \text{Var}(\tilde{z}) \equiv \sigma_z^2$. Relation (3) states that an individual would rather sacrifice p units of state 1 consumption than entertain an actuarially fair gamble impacting current *and* state 1 consumption. Proceeding as in Pratt (1964) and letting subscripts denote partial derivatives (i.e., $v_0 \equiv \partial v/\partial C_0$, $v_1 \equiv \partial v/\partial C_1$, etc.),[2]

$$(4a)\qquad v(C_0^*, C_1^* - p) = v(C_0^*, C_1^*) - v_1 p + 0(p^2)$$

---

[2] In Taylor series expansions, $0(x)$ implies $\lim_{x \to 0} 0(x)/x > 0$ and $o(x)$ implies $\lim_{x \to 0} o(x)/x = 0$.

and

$$(4b)\quad E\left[v(C_0^*+\tilde{z},C_1^*+\tilde{z})\right]$$
$$=E\left[v(C_0^*,C_1^*)+v_0\tilde{z}+v_1\tilde{z}+\tfrac{1}{2}v_{00}\tilde{z}^2+\tfrac{1}{2}v_{11}\tilde{z}^2+v_{01}\tilde{z}^2+0(\tilde{z}^3)\right]$$
$$=v(C_0^*,C_1^*)+\left(\tfrac{1}{2}v_{00}+\tfrac{1}{2}v_{11}+v_{01}\right)\sigma_z^2+o(\sigma_z^2).$$

Equating (4a) to (4b) and noting the definition (2) yields

$$(5)\quad p=\left[\frac{1}{2}\left(-\frac{U_{00}}{U_1}\right)+\frac{1}{2}\left(-\frac{U_{11}}{U_1}\right)+\left(-\frac{U_{01}}{U_1}\right)\right]\sigma_z^2.$$

Thus, while the risk premium p depends on the coefficient of absolute risk aversion, $-U_{11}/U_1$, and the term $-U_{00}/U_1$, it clearly follows from equation (5) that $\partial p/\partial(-U_{01}/U_1)>0$. Consequently, the MVRA as given in equation (1) determines the multivariate risk aversion "contribution" to the magnitude of the risk premium p.[3] And finally, as indicated in the introductory discussion, the MVRA as given in equation (1) dictates the impact of the behavioral property of multivariate risk aversion on the marginal propensity to consume.

The relationship between the behavioral attributes of multivariate risk aversion and complementarities across time, induced by nonadditive utilities, is of intuitive interest. As previously noted, multivariate risk aversion is uniquely identified by the sign of the second cross-partial derivative. The definition of complementarity in consumption preferences lacks a similar uniqueness.

In the current context, the most useful definition of complementarity is provided by Donald Katzner (1970), who notes that one possible definition of complementarity is the "well-known notion of Auspitz and Leiben that goods i and j are complements, independent, or substitutes according as the derivative $u_{ij}(x)$" is greater than, equal to, or less than zero (p. 147).[4] This definition is intuitively appealing as the decreasing marginal utility of the good is extended to its substitute via the property $\partial^2U/\partial C_0\partial C_s<0$.

The analytical focus of the above definitions implies the equivalence of multivariate risk aversion and substitutability. An intuitive interpretation may also be offered. As noted by Richard, a multivariate risk averter prefers a lottery between (high, low) and (low, high) to a lottery between (high, high) and (low, low). Thus, a multivariate risk averter will attempt to smooth his or her intertemporal consumption flows. An individual exhibiting the attribute of substitutability will also prefer the lottery (high, low),

---

[3] Richard's (1975) Theorem 3 considers two utility functions $u(x,y)$ and $v(x,y)$ satisfying 1) $u_{xy}\equiv\partial^2u/\partial x\partial y<0$, 2) $v_{xy}\equiv\partial^2v/\partial x\partial y<0$, 3) $-v_{xx}/v_x\geq-u_{xx}/u_x$, 4) $-v_{xy}/v_x\geq-u_{xy}/u_x$, 5) $-v_{yy}/v_y\geq-u_{yy}/u_y$, and 6) $-v_{xy}/v_y\geq-u_{xy}/u_y$. These conditions imply that v is more risk averse than u in both the univariate and multivariate dimensions. Thus, for $x_0<x_1$ and $y_0<y_1$,

$$u(x_1-p_x,y_0)+u(x_1,y_1-p_y)=u(x_0,y_0)+u(x_1,y_1)$$

will result in

$$v(x_1-p_x,y_0)+v(x_0,y_1-p_y)\geq v(x_0,y_0)+v(x_1,y_1).$$

That is, v would be willing to pay a greater risk premium than u to avoid the all-or-nothing gamble. Thus, while Richard similarly identifies a role for MVRA $\equiv-u_{01}/u_1$, his different methodology yielded a "global" condition encompassing univariate and multivariate risk aversion, whereas the current derivation yields the multivariate-specific comparative result $\partial p/\partial(-U_{01}/U_1)>0$.

[4] Note that the above definition of complementarity is not invariant to nonlinear positive transformations of the utility function. Nor is this definition unique. See Katzner (1970), pp. 145–50, for a critical discussion of this and alternate definitions of complementarity.

(low, high) to (high, high), (low, low).[5] In other words, the substitutability of consumption in periods 0 and 1 implies that the (high, low), (low, high) lottery assures that individual a higher level of expected utility. Thus, substitutability through time induces the individual to maximize his or her expected utility by minimizing the intertemporal variation of consumption flows; conversely, utility maximization under complementarity causes the individual to maximize the variability of intertemporal consumption choices.

II. MPC Under Arbitrary Preferences

With these preliminaries, the results are stated in Theorem 1:

THEOREM 1: *Let U be an arbitrary strictly concave utility function defined over present consumption* C_0 *and state-contingent consumption* C_s, *and exhibiting nonsatiation* w.r.t. *both arguments.*[6] *Assume that U obeys the von Neumann-Morgenstern axioms of rational choice under uncertainty, including the strong independence axiom.*

Now, consider the state-specific ratio of the coefficient of multivariate risk aversion to the coefficient of absolute risk aversion [*i.e.*, $(-U_{0i}/U_i)/(-U_{ii}/U_i)$]. *Then the marginal propensity to consume MPC for an individual exhibiting multivariate risk aversion will be negative if the weighted sum of those state-specific ratios, weighted by the* (*real*) *Arrow-Debreu state-contingent prices, exceeds unity. Further, the positivity of MPC for time-additive preferences and for multivariate risk preferrers is independent of the investment opportunity set.*

A technical explanation of Theorem 1's import is warranted here. We consider an individual at his or her consumptive-investment optimum: that is, the first-order conditions for an optimal allocation are satisfied. Now, we wish to consider the $\text{MPC} \equiv \partial C_0/\partial W_0$ of a given individual under alternative behavioral assumptions. These behavioral postulates pertain to the second cross-partial derivatives $U_{0i} \equiv \partial^2 U/\partial C_0 \partial C_i$ for a given state of the world i. If the reference individual is multivariate risk neutral, $U_{0i} \equiv 0$ for all i. If the individual is multivariate risk averse (preferrer), then $U_{0i} < (>) 0$ for all i.[7]

PROOF: We first establish the formal framework within which the proof is worked out. For the given utility function U, the consumer's constrained utility-maximization problem is

$$(6) \quad \max_{C_0,\{C_i\}} E\left[U(C_0,\tilde{C}_1)\right] \equiv \sum_{i=1}^{n} \pi_i U(C_0,C_i)$$

$$\text{s.t.}\, P_0 C_0 + \sum_{i=1}^{n} \varphi_i P_i C_i = W_0$$

for n possible states of nature, and where π_i is the probability of state i materializing at time 1, P_0 is the current price of one unit of the consumption good in terms of the numeraire ("money"),[8] φ_i is the current price of an Arrow-Debreu security paying $1 if state i materializes and 0 otherwise, and P_i is the price of the consumption good in state i.

Forming a Lagrangean in the usual fashion, the resulting first-order conditions for optimum are obtained as:

---

[5] Of course, the proof of the consumption choices exhibited under substitutability—under the mathematically equivalent property of multivariate aversion—is provided by Richard's (1975) Theorem 1, pp. 14–15.

[6] Strictly speaking, the theorem is stated in a two-period (now and then) model. However, note that C_s may be redefined as state-contingent wealth. In permitting the investor to choose state-contingent wealth, the model adopts multiperiod characteristics.

[7] Thus, throughout this paper it is assumed that $\text{sign}(U_{0i}) = \text{sign}(U_{0j})$ for all $i, j > 0$.

[8] P_0 and P_i, $i = 1, \ldots, N$, may be set identically equal to unity without loss of generality. However, the *existence* of such price indices in a multicommodity world is by no means a trivial problem. Samuelson and S. Swamy (1974) have shown that a wealth-invariant unique price index obtains only under the restrictive assumption of a homothetic utility function. Namely, if q_{ji} is the quantity of commodity j consumed in state i, we must have $C_i = f(\prod_j q_{ji}^{\alpha_j})$, where $\alpha_j, f'(\cdot) > 0$ and $\sum_j \alpha_j = 1$.

$$L = \sum_{i=1}^{n} \pi_i U(C_0, C_i) + \lambda\left(W_0 - P_0 C_0 - \sum_{i=1}^{n} \varphi_i P_i C_i\right)$$

$$\rightarrow \sum_{i=1}^{n} \pi_i \frac{\partial U(C_0, C_i)}{\partial C_0} - \lambda P_0 = 0$$

$$\pi_i \frac{\partial U(C_0, C_i)}{\partial C_i} - \lambda \varphi_i P_i = 0$$

$$W_0 - P_0 C_0 - \sum_{i=1}^{n} \varphi_i P_i C_i = 0,$$

where λ is a Lagrangean multiplier. Now, define

$$V(C_0, C_1, \dots, C_n) \equiv \sum_i \pi_i U(C_0, C_i)$$

$$p_0 \equiv P_0$$

$$p_i \equiv \varphi_i P_i \quad \text{for } i = 1, \dots, n.$$

Under this formulation,[9] the problem reduces to the consumer comparative statics analyzed in graduate microeconomics textbooks, for example, Katzner, 1970; James Henderson and Richard Quandt, 1971; and Samuelson, 1976. Thus, use the following notation:

$$\frac{\partial^2 V}{\partial C_0^2} = \sum_i \pi_i \frac{\partial^2 U}{\partial C_0^2} \equiv V_{00}$$

$$\frac{\partial^2 V}{\partial C_i^2} \equiv \pi_i \frac{\partial^2 U}{\partial C_i^2} \equiv \pi_i U_{ii} \qquad \text{for all } i,$$

$$\frac{\partial^2 V}{\partial C_0 \partial C_i} = \pi_i \frac{\partial^2 U}{\partial C_0 \partial C_i} \equiv \pi_i U_{0i} \quad \text{for all } i.$$

Now, from standard microanalysis,

$$\frac{\partial C_0}{\partial W_0} = \frac{|M_{12}|}{|M|}, \tag{7}$$

where

$$|M| = \begin{vmatrix} 0 & P_0 & \varphi_1 P_1 & \varphi_2 P_2 & \cdots & \varphi_n P_n \\ P_0 & V_{00} & \pi_1 U_{01} & \pi_2 U_{02} & \cdots & \pi_n U_{0n} \\ \varphi_1 P_1 & \pi_1 U_{01} & \pi_1 U_{11} & 0 & \cdots & 0 \\ \varphi_2 P_2 & \pi_2 U_{02} & 0 & \pi_2 U_{22} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_n P_n & \pi_n U_{0n} & 0 & 0 & \cdots & \pi_n U_{nn} \end{vmatrix} \tag{8}$$

and $|M_{12}|$ is the cofactor of the (1,2) element of M.

A solution to the RHS of (7) may be obtained from Bellman (1970). We may solve for $|M|$ and $|M_{12}|$ using the following linear algebra theorem:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D|\,|A - BD^{-1}C|, \tag{9}$$

where

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

A and D are square matrices and $|D| \neq 0$. Set

$$D = \begin{pmatrix} \pi_1 U_{11} & & 0 \\ & \pi_2 U_{22} & \\ 0 & & \pi_n U_{nn} \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & P_0 \\ P_0 & V_{00} \end{pmatrix}$$

$$B = \begin{pmatrix} \varphi_1 P_1 & \cdots & \varphi_n P_n \\ \pi_1 U_{01} & \cdots & \pi_n U_{0n} \end{pmatrix}$$

$$C = B'.$$

---

[9] Note that the strong independence axiom has implied $\partial^2 V/\partial C_i \partial C_j = 0$ for all $i \neq j, i, j > 0$. $\partial V/\partial C_i = \pi_i(\partial U/\partial C_i)$ implies $\partial^2 V/\partial C_i \partial C_j = 0$ for $i \neq j, i, j > 0$.

Then

$$|M| = \left|\begin{pmatrix} 0 & P_0 \\ P_0 & V_{00} \end{pmatrix} - \begin{pmatrix} \varphi_1 P_1 & \cdots & \varphi_n P_n \\ \pi_1 U_{01} & \cdots & \pi_n U_{0n} \end{pmatrix} \times \begin{pmatrix} \frac{1}{\pi_1 U_{11}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\pi_n U_{nn}} \end{pmatrix} \begin{pmatrix} \varphi_1 P_1 & \pi_1 U_{01} \\ \cdots & \cdots \\ \varphi_n P_n & \pi_n U_{0n} \end{pmatrix}\right|$$

preceded by the factor

$$\begin{vmatrix} \pi_1 U_{11} & & 0 \\ & \pi_2 U_{22} & \\ 0 & & \pi_n U_{nn} \end{vmatrix}.$$

We have $|D| = \prod_{i=1}^{n} \pi_i U_{ii}$. Performing the multiplication and addition $A - BD^{-1}C$, we obtain

$$|M| = -\left(\prod_{i=1}^{n} \pi_i U_{ii}\right) \times \begin{vmatrix} -\sum_i \frac{\varphi_i^2 P_i^2}{\pi_i U_{ii}} & P_0 - \sum_i \frac{\varphi_i P_i}{U_{ii}} U_{0i} \\ P_0 - \sum_i \varphi_i P_i \frac{U_{0i}}{U_{ii}} & V_{00} - \sum_i \pi_i \frac{U_{0i}^2}{U_{ii}} \end{vmatrix}$$

$$= -\left(\prod_{i=1}^{n} \pi_i U_{ii}\right)\left[\left(\sum_i \frac{\varphi_i^2 P_i^2}{\pi_i U_{ii}}\right)\left(V_{00} - \sum_i \frac{\pi_i U_{0i}^2}{U_{ii}}\right) + \left(P_0 - \sum_i \frac{\varphi_i P_i U_{0i}}{U_{ii}}\right)^2\right].$$

Note here that by the strict concavity of U,[10] $V_{00} - \Sigma_i \pi_i U_{0i}^2 / U_{ii} < 0$ and $U_{ii} < 0$. Thus, $\operatorname{sign}|M| = \operatorname{sign}(-1)^{n+1}$ as expected.

The calculation of M_{12} proceeds along similar lines. We have

$$|M_{12}| = -\begin{vmatrix} P_0 & \pi_1 U_{01} & \pi_2 U_{02} & \cdots & \pi_n U_{0n} \\ \varphi_1 P_1 & \pi_1 U_{11} & 0 & \cdots & 0 \\ \cdot & 0 & \pi_2 U_{22} & \cdots & \cdot \\ \cdot & \cdot & & \cdots & \cdot \\ \varphi_n P_n & 0 & 0 & \cdots & \pi_n U_{nn} \end{vmatrix}.$$

Thus, set

$$D_1 = D$$

$$A_1 = P_0$$

$$B_1 = (\pi_1 U_{01}, \ldots, \pi_n U_{0n})$$

$$C_1 = \begin{pmatrix} \varphi_1 P_1 \\ \cdot \\ \varphi_n P_n \end{pmatrix}.$$

Now, proceeding precisely as in the calculation of $|M|$, we have

$$|M_{12}| = -\left(\prod_{i=1}^{n} \pi_i U_{ii}\right)\left(P_0 - \sum_{i=1}^{n} \frac{\varphi_i P_i U_{0i}}{U_{ii}}\right).$$

Finally, equation (10) considers the marginal propensity to consume from (nominal)

---

[10] The strict concavity of $U(\cdot,\cdot)$ implies

$$\frac{\partial^2 U}{\partial C_0 \partial C_i}(C_0, C_i)$$

$$- \frac{\partial^2 U}{\partial C_0^2}(C_0, C_i) \frac{\partial^2 U}{\partial C_i^2}(C_0, C_i) < 0 \quad \text{for all } i.$$

Thus, $U_{0i} - U_{00} U_{ii} < 0 \rightarrow U_{0i}/U_{ii} - U_{00} > 0 \rightarrow U_{00} - U_{0i}/U_{ii} < 0$. Multiplying by π_i and summing across all i implies

$$\sum_i \pi_i U_{00} - \sum_i \pi_i U_{0i}/U_{ii} < 0$$

or

$$V_{00} - \sum_i \pi_i U_{0i}/U_{ii} < 0.$$

wealth, $\partial C_0/\partial W_0$:[11]

$$(10)\quad \frac{\partial C_0}{\partial W_0} = \frac{P_0 - \sum_i \frac{\varphi_i P_i U_{0i}}{U_{ii}}}{\left(\sum_i \frac{\varphi_i^2 P_i^2}{\pi_i U_{ii}}\right)\left(V_{00} - \sum_i \pi_i \frac{U_{0i}^2}{U_{ii}}\right) + \left(P_0 - \sum_i \frac{\varphi_i P_i U_{0i}}{U_{ii}}\right)^2}.$$

As the denominator of the RHS of (6) is positive,[12]

$$\operatorname{sign}\left(\frac{\partial C_0}{\partial W_0}\right) = \operatorname{sign}\left(1 - \sum_i \frac{\varphi_i P_i}{P_0} \frac{U_{0i}}{U_{ii}}\right) = \operatorname{sign}\left[1 - \sum_i \frac{\varphi_i P_i}{P_0}\left(-\frac{U_{0i}}{U_i}\right)\Big/\left(-\frac{U_{ii}}{U_i}\right)\right] \equiv \operatorname{sign}\left(1 - \sum_i \frac{\varphi_i P_i}{P_0} \frac{\mathrm{MVRA}_i}{\mathrm{ARA}_i}\right).$$

Thus, for a given investment opportunity set—that is, prices of (real) state-contingent claims $\{\varphi_i P_i\}$—the marginial propensity to consume will be negative only if the individual is "sufficiently" multivariate risk averse.[13]

III. Summary

Thus, differential attitudes toward multivariate risk aversion yield different results with regard to the marginal propensity to consume. Further analysis is required to investigate the results with regard to the optimal portfolio policy.[14] These additional analyses would be directed at eliciting, for both additive and nonadditive preferences, the impact of behavioral attributes on the (marginal) investment in the real/nominal riskless assets. A judicious use of first-order conditions, coupled with the well-known comparative statics techniques utilized in this paper, might yield interesting insights on these portfolio decision rules.

The general results obtained in this paper dealt explicitly with changes in consumption induced by marginal changes in nominal wealth. The transition from marginal analysis to inframarginal analysis is no simple matter. For one, we know from Gordon Pye's (1972) analysis of the time multiplicative power utility function that Theorem 1 does not simply generalize; that is, it is not possible to suppress the word "marginal" in that theorem. For another, any inframarginal analysis based on an aggregation (taking integrals)[15] of the marginal results expounded above would necessarily have to account for the fact that, under differential attitudes toward time complementarity, $\partial C_0/\partial W_0$ is evaluated at different levels of C_0 and $\{C_s\}$. It would thus require further assumptions concerning the *third*-order derivatives of the utility function. Such assumptions follow from Arrow's conjectures

---

[11] The MPC from real wealth is given by $\partial C_0/\partial(W_0/P_0)$. For constant P_0,

$$\frac{\partial C_0}{\partial(W_0/P_0)} = P_0\frac{\partial C_0}{\partial W_0} = \frac{\partial(P_0C_0)}{\partial W_0},$$

where $\partial(P_0C_0)/\partial W_0$ is the marginal consumption expenditure of nominal wealth. Naturally, since $P_0 > 0$, $\operatorname{sign}[\partial C_0/\partial(W_0/P_0)] = \operatorname{sign}(\partial C_0/\partial W_0)$.

[12] Recall from fn. 10 that

$$V_{00} - \sum_i \pi_i \frac{U_{0i}^2}{U_{ii}} \equiv \sum_i \pi_i \frac{\partial^2 U}{\partial C_0^2}(C_0, C_i) - \sum_i \pi_i \frac{U_{0i}^2}{U_{ii}} = \sum_i \pi_i \left[\frac{\partial^2 U}{\partial C_0^2}(C_0, C_i) - \frac{U_{0i}^2}{U_{ii}}\right] < 0,$$

where the negativity of $[\cdot]$ above follows from the concavity of $U(\cdot)$, which implies $U_{00} - U_{0i}^2/U_{ii} < 0$ for all $i > 0$.

[13] Note that in a world of certainty, $\pi_s = 1$, $\pi_q = 0$ for all $q \neq s$, and sign $(\partial C_0/\partial W_0)$ reduces to sign $[1 - (\mathrm{MVRA}/\mathrm{ARA})/(1+R)]$, where $1/(1+R) \equiv \varphi_s P_s/P_0$ and R is the interest rate on a real riskless bond.

[14] For example, this might be achieved by analyzing $\partial C_s/\partial W_0$ or $(\partial C_s/\partial W_0)/[1 - P_0(\partial C_0/\partial W_0)]$.

[15] That is, $C_0 \equiv \int_0^{W_0}(\partial C_0/\partial W_0)(q)\,dq$.

regarding decreasing absolute risk aversion and increasing relative risk aversion, and these postulates may yet yield additional results.

REFERENCES

Arrow, Kenneth J., "The Role of Securities in the Optimal Allocation of Risk Bearing," *Essays in the Theory of Risk Bearing*, Chicago: Markham, 1971.

Debreu, Gerard, *Theory of Value*, New York: Wiley & Sons, 1959.

Hakansson, Nils H., "Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions," *Econometrica*, September 1970, *38*, 587–607.

Henderson, James and Quandt, Richard, *Microeconomic Theory: A Mathematical Approach*, New York: McGraw-Hill, 1971.

Katzner, Donald, *Static Demand Theory*, New York: Macmillan, 1970.

Kraus, A. and Litzenberger, Robert H., "Market Equilibrium in a Multiperiod State Preference Model with Logarithmic Utility," *Journal of Finance*, December 1975, *30*, 1213–27.

Leland, Hayne E., "Saving and Uncertainty: The Precautionary Demand for Saving," *Quarterly Journal of Economics*, August 1968, *82*, 465–73.

Levhari, D. and Srinvasan, T. N., "Optimal Savings Under Uncertainty," *Review of Economic Studies*, April 1969, *36*, 153–64.

Merton, Robert C., "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, December 1971, *3*, 378–418.

______, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, September 1973, *41*, 867–87.

Mirrlees, James A., "Optimum Economic Policies under Uncertainty," unpublished manuscript, 1965.

Mirman, Leonard J., "Uncertainty and Optimal Consumption Decisions," *Econometrica*, January 1971, *39*, 179–85.

Phelps, E. S., "The Accumulation of Risky Capital: A Sequential Utility Analysis," *Econometrica*, October 1962, *30*, 729–43.

Pratt, John W., "Risk Aversion in the Small and in the Large," *Econometrica*, January-April 1964, *32*, 122–36.

Pye, Gordon, "Lifetime Portfolio Selection with Age Dependent Risk Aversion," in G. P. Szego and Shell, eds., *Mathematical Methods in Investment and Finance*, Amsterdam: North-Holland, 1972.

Richard, Scott F., "Multivariate Risk Aversion, Utility Independence and Separable Utility Functions," *Management Science*, September 1975, *22*, 12–21.

Rubinstein, Mark, "The Strong Case for the Generalized Logarithmic Utility Model as the Premier Model of Financial Markets," *Journal of Finance*, May 1976, *31*, 555–71.

Samuelson, Paul A., "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics*, August 1969, *51*, 239–46.

______, *Foundations of Economic Analysis*, New York: Atheneum, 1976.

______ **and Swamy, S.,** "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review*, September 1974, *64*, 566–93.

Farmers' Cooperatives as Behavior Toward Risk

By MILES S. KIMBALL*

The English Open Field System is important for economic history both in its own right and because the Open Field System set the stage for the subsequent Enclosure movement. One of the interesting debates about the Open Field System is over why with few exceptions each farmer held many small, widely separated, plots rather than one consolidated plot.

Donald McCloskey (1976) argues that the scattering of plots was insurance against crop failures, the idea being that if one plot did badly but another one did well, the farmer would still have enough to survive from all of his plots put together. In terms of portfolio theory, the risk to each farmer as measured by the variance of his total output could be reduced by diversifying his holdings among many small plots that each faced different weather,[1] weed, water, and soil conditions.

The main objection to McCloskey's characterization of scattering as insurance was raised by Stefano Fenoaltea (1976). He pointed out that given the costliness of having scattered plots in terms of travel time and the extra boundaries, other, cheaper modes of insurance should have been possible.[2] Although there is little evidence of formal insurance markets for crops during this period, either self-insurance through saving in good times or various informal (that is, "nonmarket") modes of insurance could have been used.

As for saving, using the numbers from McCloskey and John Nash (1984), it seems on the face of things that farmers could earn a real return on the order of 35 percent annually without the help of any bank simply by storing extra grain and selling it a few months before the next harvest when the price was at its seasonal peak. The money thus earned could then be used to purchase grain in the year after a farmer had a bad harvest. Of course, if everyone within ox or horse-cart range of the village had a bad year, the price of grain would be very high so that the money would not buy much grain, but if all the yields in the entire orbit of a village were bad, scattering of fields would not be of much help either. By whatever scheme, it was only the idiosyncratic risk of particular plots that a farmer could hope to escape, not the general risk faced by the whole village.

Suppose, however, to give McCloskey's notion of scattering as insurance the benefit of the doubt, that we assume away the possibility of self-insurance through saving (a position for which arguments will be presented in Section IV). Then as Fenoaltea points out, there still remains the possibility of an informal system of consumption loans among farmers that could perform some or all of the insurance function that formal insurance markets failed to perform. In the title, such consumption loan arrangements are labeled "farmers' cooperatives." Such mutual assistance pacts are just as likely to have appeared in the guise of the extended family, a fraternal order, or a group of a few fast friends as in the guise of something called itself a "farmers' cooperative." However, all such arrangements will be lumped together under the term "farmers' cooperatives" in the rest of the paper.

I. Modeling a Cooperative

To see how readily farmers' cooperatives could spring up and to what extent they could substitute for formal insurance markets, it is instructive to consider a simple repeated-game model of a farmers' cooperative. We will be particularly interested in 1) the conditions under which a cooperative can be self-enforcing—that is, when the temptation to default on one's obligations under the rules of the cooperative can be countered by the threat of expulsion alone. 2) the extent of the insurance a self-enforcing cooperative arrangement would be able to provide. By insisting that a cooperative be enforceable by the threat of expulsion alone, without legal or violent remedies for noncompliance, we can focus in on those informal or semiformal arrangements that could spring up with little or no institutional backing, and which might very often have escaped being recorded in any documents that would have survived down to the present.

---

*Department of Economics, University of Michigan, Ann Arbor, MI 48109. I am grateful to Jeffrey Williamson for inspiring my interest in the English Open Field System and to Gail Kimball and Jeffrey Miron for technical assistance.

[1] Hail, at least, is notorious among farmers for striking one farm without touching its neighbors.

[2] Fenoaltea's alternative explanation for the scattering of fields is that it was a way of spreading around the year each farmer's labor on his own fields as different crops (and to some extent, the same crop on different plots) needed labor at different times.

The simple repeated-game model of a farmers' cooperative is specified as follows: There are n farmers who face identical random distributions of output on their farms. The output of one farm is independent of the output of all the others. (Thus, the issue of insuring against idiosyncratic yield risks is emphasized and the issue of insuring against shocks that affect the whole community is pushed into the background.) Each year, from now into the indefinite future (that is, from $t=0$ to $t=\infty$), farmers find out how bountiful their farms will be that year (that is, they learn their yields y_t^i) and then have a choice of how much of the output of their farms they will give to other farmers and how much they will keep for themselves. It is assumed that there are no formal insurance markets available to the farmers, that farmers are unable to borrow, and that there is no carry-over or saving of grain from one harvest to the next (so that any grain kept must be consumed that year, if ever) all of which are not far from the truth for that era.[3] It is also assumed that no other method of saving is available to the farmers—a questionable assumption, but one that is appropriate here, since I am arguing that *either* farmers could save by keeping extra grain to sell a few months before harvest and then use the proceeds to buy grain in years that were bad for them *or* that voluntary cooperatives could and would have sprung up to provide insurance against idiosyncratic risk.

The cooperative itself is modeled as a set of rules for output sharing among members according to how well or how poorly each fared in a given year, with the punishment for noncompliance being expulsion from the group. It is assumed that no other cooperative would thereafter be willing to accept such a defector, so that a defector would be effectively blacklisted and forced to "go it alone" after expulsion from the cooperative.

It is best to use a very simple parameterization of utility functions and the random distribution of the yield on each plot. In line with standard practice in current macroeconomic theory, let us assume that all farmers have identical, additively time-separable, constant-relative risk aversion (and constant-intertemporal elasticity of substitution) von Neumann-Morgenstern utility functions:

$$(1)\quad U_t^i = E_t \sum_{s=0}^{\infty} \frac{u\left(c_{t+s}^i\right)}{(1+\delta)^s}$$

$$= E_t \sum_{s=0}^{\infty} \frac{1}{(1+\delta)^s}\left[\frac{\left(c_{t+s}^i\right)^{1-\rho}-1}{1-\rho}\right].$$

The operator E_t is an expectation conditional on what is known at time t. The parameter δ is the utility discount rate or "subjective rate of time preference," which here need not equal the interest rate even in the steady state, since the farmers in the model cannot borrow or save. Consumption by farmer i at date $t+s$ is denoted by c_{t+s}^i and ρ represents the Pratt-Arrow measure of relative risk aversion as well as the reciprocal of the intertemporal elasticity of substitution.

For the random annual output of each farm, let us assume a gamma distribution with the probability density function $x^{m-1}e^{-x/\beta}/\beta^m\Gamma(m)$. Besides simplifying the calculations we will need to make, this distribution meets the minimum requirement for plausibility of being bounded below by zero and has the advantage of being determined by only two parameters, a scale parameter β that will play no role in the analysis, and another parameter m which determines the "tightness" of the distribution. Since the mean of the distribution is βm and the standard deviation is $\beta\sqrt{m}$; m is related to the coefficient of variation (which measures the degree of dispersion) by $m^{-1/2}=$ coefficient of variation.

---

[3] See McCloskey and Nash (1984).

The objective of a cooperative should be to find the sharing scheme that maximizes the average welfare of the cooperative's members among the class of all schemes that can be supported by the threat of expulsion. Whether expulsion is a matter of near-indifference, or a severe punishment, will depend on the value of belonging to the cooperative. If for the moment, we take the value of belonging (V) as given, it is easy to see the general character of the optimal set of rules. Because of decreasing marginal utility, an egalitarian objective function such as average expected utility is increased if grain will be transferred from those who have more to those who have less. The optimal arrangement is for grain to be transferred from those who have more than the lowest amount to those who have the least until either 1) the quantities are equalized; or 2) each of the farmers who has more than the lowest amount has contributed as much as the threat of expulsion can force him to contribute.

Formally, the optimal sharing rule is given by the solution to the problem

$$\text{(2)}\qquad \max_{c} E_0 \sum_{t=0}^{\infty}\sum_{i=1}^{n}\frac{u(c_t^i)}{(1+\delta)^t}$$

$$\text{s.t.}\ \sum_{i=1}^{n} c_t^i \le \sum_{i=1}^{n} y_t^i\,\forall t\ \text{(physical feasibility)}$$

and

$$u(y_t^i)\le u(c_t^i)+V_t^i\,\forall i,t\ \text{(enforceability)},$$

where V is the value of belonging to the cooperative for all subsequent periods, discounted back to time t.

Since a defector is blacklisted and must fall back on his own resources,

$$\text{(3)}\qquad V_t^i = E_t\left[\sum_{s=t+1}^{\infty}\frac{u(c_s^i)}{(1+\delta)^{s-t}}\right] - E_t\left[\sum_{s=t+1}^{\infty}\frac{u(y_s^i)}{(1+\delta)^{s-t}}\right].$$

We can further simplify this; because yields are independent from one year to the next and utility is additively separable over time, there is no reason for the optimal sharing rule to depend on the past, other than to remember who is a defector. Therefore, V should be independent of time. Also, since all farmers have the same chances in the future, it seems reasonable to concentrate on symmetric sharing rules, for which the superscript i can be dropped. We can then write

$$\text{(4)}\qquad V=\frac{1}{\delta}E[u(c)-u(y)].$$

In words, V is $1/\delta$ times the unconditional expectation of the difference between a farmer's utility for a year in the cooperative and his utility outside.

II. Could a Cooperative Maintain Itself?

It is now a well-defined mathematical question when a cooperative can be self-enforcing and how much insurance it will be able to provide. A cooperative can be self-enforcing, if and only if it can pass the bootstrap test of obtaining from some severity of punishment a value of belonging to the cooperative at least equal to that severity of punishment. More formally, let us define $W(P)$ as the maximum possible worth of belonging to a cooperative that has available to it a punishment of P utils. Then V is given by

$$\text{(5)}\qquad V=\max\{P\,|\,W(P)\ge P\},$$

and $V > 0$, if and only if $W'(0) > 1$.[4] As shown in Kimball (1986), this means, for example, that a two-farmer cooperative—in which such a bootstrap operation of enforcing cooperation by the value of cooperation itself is the hardest to get going[5]—can be self-enforcing, if and only if

$$\frac{\int_0^{1/2} x^{m-\rho-1}(1-x)^{m+\rho-1}dx}{\int_0^1 x^{m-1}(1-x)^{m-1}} - \frac{1}{2} \geq \delta. \tag{6}$$

The numerator and denominator are incomplete and complete beta functions, respectively, which can be found in standard tables.[6] The value of the entire left-hand side of the inequality above, which shows the maximum utility discount rate at which a two-farmer cooperative can be sustained, can be seen listed in Table 1 for various parameter values. The box indicates the range of parameter values I will argue to be most plausible.

The appropriate range of values for m can be ascertained from the coefficients of variation of plot yields given by McCloskey (1976). For a single plot and a mix of different grains, McCloskey estimates a coefficient of variation of .44, which corresponds to a value of 5.17 for m. For different grains individually and averages over different groups of manors, the highest coefficient of variation listed is .64 for oats in the Woodstock Manors, which corresponds to a value of 2.44 for m, and the lowest coefficient of variation listed is .35 for barley in the Winchester Manors, which corresponds to a value of 8.16 for m. Therefore we will be relatively safe if we allow a range for m of 3 to 8 around the preferred value of 5.

TABLE 1—UTILITY DISCOUNT RATE BELOW WHICH A TWO-FARMER COOPERATIVE CAN BE MAINTAINED (Percent per Year)

	Coefficient of	ρ					
m	Variation	$\frac{1}{2}$	1	2	3	4	6
1	1.00	78.5	∞	∞	∞	∞	∞
2	.71	33.9	107.0	∞	∞	∞	∞
3	.58	24.0	71.9	531.3	∞	∞	∞
4	.50	19.3	53.1	262.5	1930.0	∞	∞
5	.45	16.5	43.3	177.5	824.8	6936.1	∞
6	.41	14.6	37.1	136.2	491.7	2455.2	∞
7	.38	13.2	32.8	110.7	350.6	1334.3	92338.7
8	.35	12.2	29.6	95.6	272.5	876.3	24463.1
9	.33	11.3	27.1	84.0	223.5	639.8	10549.0
10	.32	10.6	25.1	75.3	190.0	499.4	5895.2
16	.25	8.0	18.3	48.9	104.1	212.7	1018.2
25	.20	6.2	13.8	34.3	66.2	118.3	372.3

There is one necessary qualification to our use of McCloskey's figure for the coefficient of variation. His figure of .44 is for the total variation in output, including the component of variability shared by the whole village, while the model we have constructed focuses on the idiosyncratic, plot-to-plot yield variation alone. This suggests using somewhat lower values of the coefficient of variation (and higher values of m) to represent this idiosyncratic portion of yield variation. When adjusting for this factor, however, it should be recognized that the figures in Table 1 (though not those in Table 3, below) are conservative because they are limited to two-farmer cooperatives, which are less able to pull themselves up by their own bootstraps than a many-farmer cooperative, and that with anything like constant relative risk aversion, being subjected to one set of risks (for example, the aggregate uncertainty faced by the village) makes one act more averse to other independent risks[7] (for example, the idiosyncratic yield risk of a given plot).

---

[4] It is clear because of strictly diminishing marginal utility that $W(P)$ is strictly concave, so that it will always remain below the 45° line for positive P if $W'(P) \leq 1$. Conversely, if $W'(0) > 1$, the graph of $W(P)$ must be above the 45° line for some strictly positive values of P.

[5] The welfare gains from sharing that can be obtained with a given size of punishment will go up as more farmers are added because the transfers will on average be to more destitute farmers; therefore the $W(P)$ function must shift upward as more farmers are added. This upward shift of the $W(P)$ function will make it more likely that some of the positive part of the 45° line will be underneath the graph of $W(P)$. Once some portion of the $W(P)$ graph is above the 45° line, an upward shift will raise the upper intersection of $W(P)$ with the 45° line (which is at (V, V)). Thus, once the cooperative is running, adding more farmers to the system will increase the expected utility of each and the dread each has of expulsion.

[6] See, for example, Karl Pearson (1934).

Unfortunately, we have relatively little guidance in guessing the initial coefficient of relative risk aversion for the typical medieval farmer. Even the relative risk aversion of modern consumers and investors is still very much in doubt. To give some idea of the range of estimates, N. Gregory Mankiw, Julio Rotemberg, and Lawrence Summers (1985) report values of ρ between zero and one, Lars Hansen and Kenneth Singleton (1983) report values between zero and two, Mankiw (1985) reports values between two and four, and Robert Hall (1981) reports values around 15. Among authors offering less formal estimates, Irwin Friend and Marshall Blume (1975) suggest a value above two, Sanford Grossman and Robert Shiller (1981) recommend a value of four, and Friend and Joel Hasbrouck (1982) suggest a value of six.

Faced with this bewildering array of estimates, perhaps the best procedure is to provide a table that will allow one to use introspection and casual observation to judge one's own risk aversion and that of one's contemporaries and then to add to that number according to how much more risk averse one believes medieval English farmers to have been. Therefore, Table 2 indicates how great a risk one would be willing to take with one's permanent income given various coefficients of relative risk aversion. Imagining that one has an initial income and consumption of \$50,000 guaranteed for the rest of one's life, one is asked to consider a bet that has a 50-50 chance of either doubling one's guaranteed yearly income and consumption to \$100,000 per year for the remainder of one's life or reducing one's yearly income and consumption to $\$x$ per year for the rest of one's life. The table shows the value of x for which individuals with various degrees of risk aversion would be indifferent to such a bet. Table 2 should convince most readers that .5 is too low a value for ρ and that 10 is too high a value for ρ, even given a modest addition to ρ due to the presumed higher-risk aversion of medieval farmers.[8] But even with the help of introspection, values of ρ anywhere from 1 to 6 seem difficult to rule out given the present state of our knowledge.

TABLE 2—EVEN BETS THAT ARE EQUIVALENT TO \$50,000

ρ	FOR SOMEONE WITH A RELATIVE RISK AVERSION OF ρ, A LIFELONG ANNUAL INCOME OF \$50,000 PER YEAR IS EQUIVALENT TO A 50-50 CHANCE OF A LIFELONG ANNUAL INCOME OF		
0	\$100,000	or	\$ 0
$\frac{1}{2}$	\$100,000	or	\$17,157
1	\$100,000	or	\$25,000
2	\$100,000	or	\$33,333
3	\$100,000	or	\$37,976
4	\$100,000	or	\$40,548
6	\$100,000	or	\$43,665
10	\$100,000	or	\$46,299
20	\$100,000	or	\$48,209

While we are considering the likely range of values for ρ, it is interesting to note that McCloskey's claim that risk aversion was the motivation behind scattering would imply a relative risk aversion somewhere above two. McCloskey (1976) estimates that scattering reduced the coefficient of variation of net output from .440 to .347 and that this was worth an estimated 10 percent reduction in mean output due to scattering. Using the same assumptions about functional form we have used all along (which admittedly, are somewhat different from those used by McCloskey), it is possible to calculate the utility gain from the reduction in coefficient of variation from .440 to .347 in terms of the equivalent percentage change in mean output. A farmer who had a relative risk aversion of two would be willing to accept only an 8.3 percent reduction in the scale of output for the sake of reducing the coefficient of variation from .440 to .347; therefore, according to McCloskey's estimates of costs and benefits, higher levels of risk aversion would be needed to motivate scattering. (A relative risk aversion of 2.5 is high enough, since it would make the reduction in coefficient of variation from .440 to .347 worth an 11 percent reduction in the scale of output.)

---

[7] See John Pratt and Richard Zeckhauser (1987).

[8] In judging how much more risk averse medieval farmers were than we are, it is important to remember that while halving annual income would hurt them much more than it would us, doubling annual income would be a much greater boon to them than to us. Thus it is not clear how much greater their *relative* risk aversion was than ours.

Having determined reasonable ranges for the parameters m and ρ, one can see in Table 1 that within these ranges all but two of the lower left-hand entries within those ranges are above even the 35 percent annual medieval interest rate estimated by McCloskey and Nash (1984), which I will argue in Section IV is a serious overestimate of the utility discount rate δ. Therefore, it appears that farmers' cooperatives of the sort we are considering would have been quite viable.

III. How Much Could a Cooperative Do?

If, then, we accept the proposition that informal cooperatives could be held together by the threat of expulsion alone, we are left with the issue of how much a cooperative could require its members to do without tempting them to defect or drop out at some crucial moment. Would a cooperative have to limit its activities to the occasional contribution of a few bushels of grain to a farmer whose crop was totally wiped out, or could it enforce a communitarian ideal of full sharing of all output?

By continuity, we know that a cooperative could not force its members to do very much if the utility discount rate was too close to the bootstrap-cutoff level, which, for two-farmer cooperatives, is indicated in Table 1. At lower utility discount rates, a cooperative could enforce more and more sharing of output, until the cooperative approached the ideal of full sharing.

As it turns out, full sharing could have been achieved at positive discount rates only if relative risk aversion were equal to one. It is shown in Kimball (1986) by example that when $\rho < 1$, diminishing marginal utility is weak enough that a very fortunate farmer enjoys feasting on his hoard so much that he cannot be induced to share it all. On the other hand, when $\rho > 1$, poverty is so painful that when all farmers in the cooperative are doing badly, but one is doing a little better than the others, he cannot be induced to share all of his pitiful pile. Still, the special case of logarithmic utility when $\rho = 1$ can provide a rough but useful indication of how far the utility discount rate has to be below the bootstrap-cutoff level for a cooperative to engage in extensive output sharing. Table 3 shows the maximum utility discount rates at which full sharing could be enforced in an n-farmer cooperative made up of farmers who have a relative risk aversion of one.[9]

TABLE 3—MAXIMUM δ CONSISTENT WITH FULL SHARING WHEN $\rho = 1$ (PERCENT PER YEAR)

m	Coefficient of Variation	n = 2	3	5	10	20
1	1.00	44.27	36.54	29.44	22.86	18.43
2	.71	20.22	16.81	13.64	10.65	8.61
3	.58	13.01	10.85	8.83	6.91	5.59
4	.50	9.58	8.00	6.52	5.11	4.14
5	.45	7.57	6.34	5.17	4.05	3.28
6	.41	6.26	5.24	4.28	3.36	2.89
7	.38	5.34	4.47	3.65	2.86	2.32
8	.35	4.65	3.90	3.18	2.50	2.03
9	.33	4.12	3.45	2.82	2.22	1.80
10	.32	3.70	3.10	2.53	1.99	1.61
16	.25	2.29	1.92	1.57	1.24	1.00
25	.20	1.46	1.22	1.00	.79	.64

One of the first things one notices about Table 3 is that while I have argued that it is easier to get a cooperative going with more farmers, it is harder to always enforce full sharing in a larger group, since occasionally, one well-off farmer among destitute fellows would be required to split his grain n ways. Still, Table 3 suggests that even such extreme feats of sharing can often be enforced at the utility discount rates of a few percent ordinarily assumed for modern households. Since for a relative risk aversion of one, 35 percent is near or even above the bootstrap-cutoff level, it makes a big difference, at least in this special case, whether one believes medieval English farmers had utility discount rates similar to ours or had utility discount rates close to the market interest rate that prevailed in that era.

---

[9] The calculation involved is straightforward, though it involves the use of tables of the psi function $\psi(x) = \Gamma'(x)/\Gamma(x)$ such as those in Harold Davis (1933). The derivation is given in more detail in Kimball (1986).

IV. How Impatient Were Medieval English Peasants?

For many reasons, it is hard to believe that English farmers a few centuries ago were so much more impatient than we are. As McCloskey and Nash (1984) point out, "Peasants worked hard at ploughing fallow land and cleaning ditches in aid of yields quite far removed" (p. 185), which evidences considerable patience and thought for the future. Also, men in the Middle Ages have the reputation for being very concerned with the afterlife, which was not so much nearer to them nor so much weightier a matter to them than to, say, modern fundamentalists, that a 35 percent utility discount rate would not have substantially vitiated their concern. Finally, utility discount rates high enough to prevent cooperation in a case in which there was so much to gain as in the sharing of grain should have undermined cooperation among neighbors in many lesser matters as well, in a way that—along with the reduced concern for the future welfare of children that a high discount rate would be likely to induce—would have made medieval peasants seem considerably nastier than is evidenced in the historical record.

If medieval farmers were not much more impatient than we, though, how could the market interest rate remain so high? McCloskey and Nash put the question well:

> If interest rates were so high, why were the Europeans so poor? A few years of moderate abstention invested at over 20 percent would make a man rich. True, men of all sorts, bond and free, did in fact scale the social ladder quickly in the Middle Ages.... But the point is that the whole society could and would have done so. [1984, p. 185]

There are two possible solutions to this puzzle. One is that the upper classes dissipated their savings by war, politics, and other varieties of rent seeking that might have a high private return but a low social return, while the lower classes used up most of what would have been savings in keeping relatives and friends from starvation in exactly the kind of cooperative described here. In helping the destitute, a somewhat better-off peasant would exhibit considerable patience, even though the destitute themselves understandably would have been willing to sell their birthright for a mess of potage if they had not been able to get help from their neighbors or relatives.

The other possible solution to the puzzle is that the gentry, being near monopolists in the local loan market (and perhaps in the local off-season grain market), were reluctant to save and lend out more, lest they drive their rate of return down, while the peasants were simply unable to protect any substantial amount of savings. As McCloskey puts it:

> The plow team and other movable goods a peasant might have [including, presumably, money or exceptionally large stores of grain], were in any case hazardous themselves, subject as they were to seizure and taxation.... The lord of the manor could follow such advice [to save money], rich and secure as he was, without his peasants being in a position to do so as well.
> [1976, p. 163]

Such circumstances could have made the capital market imperfect enough to drive the market rate of interest and the utility discount rate far apart and make our assumption of no saving by peasants of money or grain (other than seed grain) a forgivable simplification.

V. Conclusion

If utility discount rates in the Middle Ages were anything like they are now—if medieval peasants were perhaps a trifle more impatient because they were poorer and also discounted the future a few percent more per year because of their higher adult mortality rates—then it can be seen from Table 3 that a farmers' cooperative could have provided a substantial amount of insurance even among farmers with a degree of risk aversion ($\rho = 1$) that would not justify scattering. Farmers with higher levels of risk aversion ($\rho > 1$) could have formed self-enforcing cooperatives involving substantial sharing of output even at very high discount rates.[10] These cooperatives could not enforce full sharing in very bad times, but in normal times would provide insurance at a much lower cost than scattering of plots. It is possible that farmers were so risk averse that they would diversify their land holdings merely for the sake of bad years when full sharing could not be enforced, but this is a much more doubtful proposition than that they would use scattering in the absence of any other good mode of insurance. It is difficult to determine exactly how great a limitation on the effectiveness of cooperatives it would be that farmers valued their grain too much in times of general dearth. However, if there were many farmers in the cooperative, to whom membership was worth a great deal, this would cause problems only a small fraction of the time—too seldom to make farmers fall back on scattering for the purpose of insurance.

With all of the foregoing arguments, I do not mean to claim that farmers' cooperatives would have worked well in every village. Many factors could affect the success of a cooperative. On the positive side, the altruism that goes with blood ties and genuine friendship, or any noneconomic value of belonging to the cooperative (for instance, being invited to parties or being greeted in the way) would make sharing arrangements work better. On the negative side, quarrelsomeness, the desire to trump up pretexts to dump currently destitute members from the cooperative, or too great a readiness to forgive renegades would tend to undermine sharing. On average, the factors aiding cooperation should have been as strong as those hindering it. However, in particular villages, the negative factors may have won out. In some cases it might be possible to identify factors that hindered sharing arrangements in certain regions and thereby led to an increased dependence on scattering as insurance.

It is also *possible* that some crucial obstacle can be found that prevented effective sharing arrangements in virtually all of Medieval England, but at a minimum, an explanation for the failure of cooperative insurance must be added to the explanation of the failure of market insurance by those who claim that the scattering of plots in the Open Field System persisted primarily as a mode of insurance.

REFERENCES

Davis, Harold T., *Tables of the Higher Mathematical Functions*, Bloomington: Principia Press, 1933.

Fenoaltea, Stefano, "Risk, Transaction Costs, and the Organization of Medieval Agriculture," *Explorations in Economic History*, April 1976, *13*, 129–51.

Friend, Irwin and Hasbrouck, Joel, "The Effect of Inflation on the Profitability and Valuation of U.S. Corporations," in M. Sarnat and G. Szego, eds. *Savings, Investment and Capital Markets in an Inflationary Economy*, Cambridge: Ballinger, 1982.

______, **and Blume, Marshall E.**, "The Demand for Risky Assets," *American Economic Review*, December 1975, *65*, 900–22.

Grossman, Sanford J. and Shiller, Robert J., "The Determinants of the Variability of Stock Prices," *American Economic Review*, May 1981, *71*, 222–27.

Hall, Robert E., "Intertemporal Substitution in Consumption," National Bureau of Economic Research Working Paper No. 720, Cambridge, 1981.

Hansen, Lars Peter and Singleton, Kenneth J., "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, April 1983, *91*, 249–65.

Kimball, Miles S., "Farmers' Cooperatives as Behavior Towards Risk," unpublished working paper, 1986.

---

[10] Just how substantial this sharing would be can be guessed by extrapolation from Tables 1 and 3 and could with some effort be calculated through computer simulations, which I will not attempt here.

Mankiw, N. Gregory, "Consumer Durables and the Real Interest Rate," *Review of Economics and Statistics*, August 1985, *67*, 353–62.

______, **Rotemberg, Julio J. and Summers, Lawrence H.,** "Intertemporal Substitution in Macroeconomics," *Quarterly Journal of Economics*, February 1985, *100*, 225–51.

McCloskey, Donald N., "English Open Fields as Behavior Towards Risk," in P. Uselding, ed., *Research in Economic History* 1976, *1*, 124–70.

______ **and Nash, John,** "Corn at Interest: The Extent and Cost of Grain Storage in Medieval England," *American Economic Review*, March 1984, *74*, 174–87.

Pearson, Karl, *Tables of the Incomplete Beta Function*, Cambridge, England: University Press, 1934.

Pratt, John W. and Zeckhauser, Richard J., "Proper Risk Aversion," *Econometrica*, January 1987, *55*, 143–54.

Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?

By UZI SEGAL*

One of the most puzzling paradoxes in decision theory is the preference reversal phenomenon. This phenomenon seems to contradict the transitivity axiom, for a long time, one of the cornerstones of utility theory. This paradox is established whenever a decision maker prefers lottery X to lottery Y, but is willing to put a lower selling price on X than on Y. Such experiments, first reported by Harold Lindman, 1971; and Sarah Lichtenstein and Paul Slovic, 1971, were repeated by David Grether and Charles Plott, 1979; Werner Pommerehne et al., 1982; and Robert Reilly, 1982. Although the later researchers improved the mechanism through which the selling price emerges, they all found systematic reversals.

A new approach to this problem was developed by Charles Holt, 1986, and Edi Karni and Zvi Safra, 1987. From two different starting points, these authors showed that the preference reversal phenomenon does not necessarily prove a violation of the transitivity axiom, as it may contradict the independence axiom. This implies that people do not maximize expected utility, but in that case the preference reversal phenomenon becomes just another evidence against expected utility theory, but not against the transitivity axiom.

These results have their own disadvantages. Although the independence axiom is not as fundamental as transitivity, it is nevertheless very appealing on normative grounds. Of course, the preference reversal phenomenon necessarily contradicts at least one of the assumptions of expected utility theory, but empirical evidence shows that despite its normative appeal, violations of the reduction of compound lotteries axiom may happen (see, for example, Joshua Ronen, 1971, and Doug Snowball and Clif Brown, 1979). In this paper I therefore suggest a decision mechanism for the preference reversals lotteries which is transitive and satisfies the independence axiom, but not the reduction axiom.

I. The Preference Reversal Phenomenon

Consider the two lotteries $P=(-1,1/36;\ 4,35/36)$ and $S=(-1.5,25/36;\ 16,\ 11/36)$.[1] P is sometimes called the P bet (or lottery) and S is called the \$ lottery. After the decision maker expresses his preferences between these two lotteries, he is asked to announce the prices for which he is willing to sell P and S. This is done by using the Gordon Becker, M. DeGroot, and J. Marschak (1964) mechanism. After the decision maker declares his selling price of P, call it r_p, one number out of the set $(0.00,\dots,9.99)$ is selected at random (the offer price). If it exceeds r_P, the decision maker receives the offer price. Otherwise, he participates in the lottery P. Similarly, the selling price of S, r_S, is found. Denote these new compound lotteries by X_P and X_S, respectively. This mechanism seems to force the decision maker to announce his true certainty equivalent[2] of P and S. Indeed, if $r_P > CE(P)$, he may have to participate in P even though he would rather sell it, while $r_P < CE(P)$ may force him to sell P against his will. If $r_P = CE(P)$ and $r_S = CE(S)$, then by the definition of the certainty equivalent and by the transitivity axiom, $P \succsim S$[3] iff $r_P \geq r_S$. However, most people prefer P to S, but announce r_P, which is less than r_S.

---

*Department of Economics, University of Toronto, Toronto, Canada M5S 1A1. I wrote this paper while visiting the Economics Department at UCLA. I am grateful to Edi Karni, Mark Machina, Zvi Safra, and an anonymous referee for their comments.

[1] The lottery $(x_1, p_1;\dots; x_n, p_n)$ yields x_i dollars with probability p_i, $i=1,\dots,n$.

[2] The certainty equivalent of a lottery A is given implicitly by $(CE(A),1)\sim A$.

Before presenting other explanations for this phenomenon, one should define precisely the independence and reduction axioms. By the later one, the decision maker is indifferent between a compound lottery, that is, a lottery at which the prizes are by themselves tickets for other (standard) lotteries, and its actuarially equivalent one-stage lottery. By the independence axiom, the simple lottery X is preferred to the simple lottery Y, iff the compound lottery A, yielding with probability p a ticket for X and with probability $1-p$ a ticket for another lottery Z is preferred to B, which is the same as A but with Y instead of X. These two axioms imply together that the decision maker prefers the simple lottery $(x_1, p_1; \dots; x_n, p_n)$ to $(y_1, q_1; \dots; y_m, q_m)$, iff for every $0 < p \leq 1$ he prefers the simple lottery $(x_1, pp_1; \dots; x_n, pp_n; z_1, (1-p)r_1; \dots; z_l, (1-p)r_i)$ to $(y_1, pq_1; \dots; y_m, pq_m; z_1, (1-p)r_1; \dots; z_l, (1-p)r_l)$. Most recent writers refer (at least implicitly) to this last axiom as the independence axiom. However, in this paper the independence preference axiom is to be understood in its two-stage lotteries context.

Two recent articles showed that the preference reversal phenomenon, although inconsistent with expected utility theory, could alternatively be attributed to a violation of the independence axiom and not the transitivity axiom. Holt (1986) pointed out that the experiment itself might have some influence on the announced selling prices. Indeed, to save money, the subjects participated only in one lottery. The decision maker first decided whether he prefers P or S (say, $P \succ S$), and announced r_P and r_S. Then, one out of the lotteries P, X_P, and X_S was selected at random, and the decision maker played it. Formally, he had to choose between $A = (P, 1/3; X_P, 1/3; X_S, 1/3)$ and $B = (S, 1/3; X_P, 1/3; X_S, 1/3)$. By the independence axiom $P \succ S$, iff $A \succ B$. However, if the independence axiom is violated, the decision maker may prefer S to P (and announce $r_S > r_P$), but prefer A to B.

Karni and Safra (1987) took a different approach. They proved that even if the decision maker is to participate in all three lotteries, he may still prefer P to S with $r_S > r_P$. According to these authors, X_P and X_S should be considered two-stage lotteries. Suppose, for example, that $r_S = 3.50$. In that case, $X_S = (S, 0.351; \; 3.51, \; 1/1000; \dots; 9.99, 1/1000)$. In the general case, $X_S = (S, 0.1(r_S + 0.01); r_S + 0.01, 1/1000; \dots; 9.99, 1/1000)$. By the independence axiom, $X_S \sim (CE(S), 0.1(r_S + 0.01); r_S + 0.01, 1/1000; \dots; 9.99, 1/1000)$, and the decision-maker's optimal strategy is to announce $r_S = CE(S)$. This analysis crucially depends on the independence axiom. However, if the decision maker does not satisfy this axiom, there is no reason to assume that $r_S = CE(S)$. Karni and Safra proved, by means of a non-expected utility example, that if the decision maker reduces compound lotteries to simple lotteries by calculating the corresponding probabilities, (i.e., if he accepts the reduction axiom), then he may prefer P to S with $r_S > r_P$. Moreover, they proved that for all non-expected utility theories, preference reversal-type behavioral patterns necessarily exist.

These two papers claim that the preference reversal phenomenon does not necessarily imply nontransitive preferences, as decision makers may disobey the independence axiom. Although this axiom is not essential for decision theory, it is very acceptable on a normative ground and it is not clear that decision makers are not willing to accept it as a decision rule. The empirical literature concerning this question present opposing results. Some consider the Allais paradox as a violation of this axiom (Mark Machina, 1982), but other empirical results indicate acceptance of this axiom, at least at its two-stage lotteries version (Daniel Kahneman and Amos Tversky, 1979), and a rejection of the reduction axiom (Ronen, 1971, and Snowball and Brown, 1979; see also Uzi Segal, 1984). As I show in Section II, the preference reversal phenomenon may prove a violation of the reduction axiom rather than the independence axiom.

---

[3] $P \succsim S$ iff P is weakly preferred to S. $P \succ S$ iff $P \succsim S$, but not $S \succsim P$, and $P \sim S$ iff $P \succsim S$ and $S \succsim P$.

II. Preference Reversals and the Reduction Axiom

Consider again the lottery X_S. In this lottery the decision maker announces a number r_S. Then he is offered a random price from the set $(0.00,\dots,9.99)$. If it exceeds r_S he receives this amount, otherwise he plays the lottery S. Karni and Safra claimed that this lottery gives with probability $0.1(r_S+0.01)$ the lottery S, and with probability $1/1000$ each of the prizes $r_S+0.01,\dots,9.99$. However, other possible interpretations lead to different results. Let $\langle\alpha,\beta\rangle$ be the uniform distribution on the segment $[\alpha,\beta]$. It may be that the decision maker considers X_S as the following two-stage lottery. With probability $0.1(r_S+0.01)$ it gives him a ticket to the lottery S. On the other hand, if the offer price exceeds r_S, the probability of which is $0.1(9.99-r_S)$, then he participates in the lottery $\langle r_S+0.01,9.99\rangle$ which is the uniform distribution over the segment $[r_S+0.01,9.99]$. In other words, $X_S=(S,0.1(r_S+0.01);\langle r_S+0.01,9.99\rangle,0.1(9.99-r_S))$. By using the independence axiom and the certainty equivalent mechanism one obtains that

$$(1)\quad X_S\sim\big(CE(S),0.1(r_S+0.01);$$
$$CE(\langle r_S+0.01,9.99\rangle),0.1(9.99-r_S)\big).$$

Similarly,

$$(2)\quad X_P\sim\big(CE(P),0.1(r_P+0.01);$$
$$CE(\langle r_P+0.01,9.99\rangle),0.1(9.99-r_P)\big).$$

It may now happen that $P\succ S$, but $r_S>r_P$. This is demonstrated by the following example.

Example: Let A be a lottery. Define $F_A(x)=\Pr(A\le x)$, and let $A^0=\{(x,p)\in\mathbf{R}\times[0.1]: F_A(x)\le p\}$. Let $L=\{A: F_A(-2)=0\}$. L is the set of all lotteries with loss floor of -2. For every $A,B\in L$, let $A\succsim B$ iff $V(A)\ge V(B)$, where

TABLE 1—VALUES OF X_P AND X_S

r_P	$V(X_P)$	r_S	$V(X_S)$
3.80	5.73982	3.80	5.73260
3.81	5.73988	3.81	5.73265
3.82	5.73993	3.82	5.73268
3.83	5.73804	3.83	5.732705
3.84	5.73807	3.84	5.732719
3.85	5.73808	3.85	5.732723
3.86	5.738093	3.86	5.732718
3.87	5.738091	3.87	5.73270
3.88	5.73808	3.88	5.73268
3.89	5.73806	3.89	5.73264

$$(3)\quad V(A)=\mu(A^0\cap[-2,\infty)\times[0,1])$$
$$+1.44\cdot10^8\mu\left(A^0\cap\left[\frac{555}{144},\frac{556}{144}\right]\right.$$
$$\left.\times\left[0,\frac{1}{1000}\right]\right)\mu\cdot\left(A^0\cap\left[\frac{936}{144},\frac{937}{144}\right]\right.$$
$$\left.\times\left[\frac{1380}{3600},\frac{1381}{3600}\right]\right).$$

(μ denotes the Lebesgue's measure.) It is easy to verify that $CE(P)=278/72$ and $CE(S)=277/72$, hence $P>S$. Similarly, for $3.03\le x\le 4.32$,

$$CE(\langle x,9.99\rangle)=\frac{x+9.99}{2}-\frac{5}{2592}.$$

From (1), (2), and (3) we now obtain Table 1. It follows that $r_P=3.82$, $r_S=3.85$, and $r_P<r_S$.

Remark 1: The preference relation in this example is transitive and continuous, satisfies the first-order stochastic dominance axiom for preferences[4] and the independence axiom, but not the reduction of compound lotteries axiom.

---

[4] First-order stochastic dominance axiom for preferences: If, for every x, $F_A(x)\le F_B(x)$, then $A\succcurlyeq B$.

Remark 2: V is not a measure of A^0. It can be proved that if V is a measure, then this analysis will not solve the preference reversal phenomenon.

III. Concluding Remarks

Holt (1986) and Karni and Safra (1987) claimed that the preference reversal phenomenon does not necessarily contradict the transitivity axiom, as it may indicate a violation of the independence axiom. In this paper I showed that this phenomenon may suggest a departure from the reduction of compound lotteries axiom, and not from the independence axiom. What is missing in this approach, as well as in the above-mentioned two works, is a general framework that explains the preference reversal phenomenon together with other paradoxes like the Allais paradox and the common ratio effect. First steps in this direction were taken by Holt (1986).

REFERENCES

Becker, Gordon M., DeGroot, Morris H. and Marschak, Jacob, "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, July 1964, *9*, 226–32.

Grether, David M. and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, September 1979, *69*, 623–38.

Holt, Charles A., "Preference Reversals and the Independence Axiom," *American Economic Review*, June 1986, *76*, 508–15.

Kahneman, Daniel and Tversky, Amos, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, March 1979, *47*, 263–91.

Karni, Edi and Safra, Zvi, "Preference Reversals and the Theory of Choice Under Risk," *Econometrica*, July 1987, *55*, 675–85.

Lichtenstein, Sarah and Slovic, Paul, "Reversals of Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, July 1971, *89*, 46–55.

Lindman, Harold R., "Inconsistent Preferences Among Gambles," *Journal of Experimental Psychology*, May 1971, *89*, 390–97.

Machina, Mark J., "'Expected Utility' Analysis Without the Independence Axiom," *Econometrica*, March 1982, *50*, 277–323.

Pommerehne, Werner W., Schneider, Friedrich and Zweifel, Peter, "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination," *American Economic Review*, June 1982, *72*, 569–74.

Reilly, Robert J., "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design," *American Economic Review*, June 1982, *72*, 576–84.

Ronen, Joshua, "Some Effects of Sequential Aggregation in Accounting and Decision-Making," *Journal of Accounting Research*, Autumn 1971, *9*, 307–32.

Segal, Uzi, "Nonlinear Decision Weights with the Independence Axiom," Working Paper No. 353, UCLA, 1984.

Snowball, Doug and Brown, Clif, "Decision Making Involving Sequential Events: Some Effects of Disaggregated Data and Dispositions Towards Risk," *Decision Sciences*, 1979, *10*, 527–46.

Uncertain Tax Policies, Individual Behavior, and Welfare

By JAMES ALM*

> Certainty is illusion, and repose is not the nature of man.
>
> [Oliver Wendell Holmes, Jr., *The Path of the Law*]

A common complaint about government is that frequent, seemingly random changes in its policies make economic decisions risky. Uncertainty about income tax policies is often cited as a particular problem for individuals. In the 1960s and 1970s there were nearly a dozen significant changes in the tax laws, and in this decade alone there has been major tax action every year. The recent passage of federal tax reform reduces some immediate uncertainties, but it is widely believed that the reformed system will eventually be revised, in ways that are not now known. Difficulties in interpreting the existing tax laws also introduce uncertainty. And tax policies do not in fact have to change to make individuals uncertain. Merely the discussion in Congress of potential tax changes introduces some element of risk into individual planning.

Uncertain tax policies make individual choices in a variety of dimensions more difficult. Individuals who are planning their financial affairs would like to know whether their itemized deductions will be allowed. They also need to know whether tax changes will alter the return on their tax shelters by changing such things as depreciation rules, investment tax credits, interest deductions, at-risk rules, or capital gains tax rates and holding periods. Potential taxation of employee benefits makes worker choice between wage and nonwage compensation risky. Actual and proposed changes in federal tax policies toward saving—the tax-exempt status of interest on state and local debt, the deductibility of contributions to retirement programs, the treatment of estate and gift taxes, the preferential treatment of capital gains, to name but a few—have increased the riskiness to saving. Individuals do not know whether they will be audited by the authorities or how much unreported income will be discovered by government auditors. The tax rate schedule itself has been altered numerous times in the past. Indeed, uncertain income tax policies affect virtually the entire range of individual choices: to work, to save, to invest, to spend, to receive compensation, and to evade.

There is an enormous theoretical and empirical literature that analyzes how individuals respond to known, certain tax and transfer policies.[1] There has also been much work on how individuals behave in the presence of uncertainty.[2] However, there has been little analysis of how individuals act when it is tax policy that is both present and uncertain.[3]

The purpose of this paper is to examine how individuals respond to greater uncertainty about individual income tax policies. Two basic types of risk are analyzed. The first is called "tax base risk." Here the individual does not know with certainty whether the government will change the basic nature of the tax base. A second type of risk is called "tax rate risk," in which the tax rate applied to the (certain) base becomes riskier.

---

*Department of Economics, University of Colorado, Boulder, CO 80309-0256. I would like to thank Jonathan Skinner and several referees for helpful comments.

[1] Much of this literature is summarized in Sheldon Danziger, Robert Haveman, and Robert Plotnick (1981).

[2] For a discussion of this literature, see John Hey (1979).

[3] Lawrence Weiss (1976) and Joseph Stiglitz (1982) examine this issue indirectly. They determine the conditions in a one-consumer economy under which random taxation in the labor supply decision leads to a Pareto improvement. F. R. Chang and D. E. Wildasin (1986) examine the optimality of local randomization of commodity taxes. Yves Balcer and Efraim Sadka (1982) examine the issue of randomization in a many-consumer economy. More recently, Jonathan Skinner (1986) estimates the welfare cost of uncertain taxation to be $12 billion in 1985.

The difference between base and rate risk is not always clear-cut, and in some cases becomes a matter of interpretation; still, it is often a natural and intuitive distinction. In each case an increase in risk is measured as a mean-preserving spread in the distribution of the random variable. The principal conclusion is that increased risk has substantial allocative effects on the decisions to invest in tax shelters, to receive compensation in wage or nonwage forms, to spend on tax deductible items, to save, to underreport one's income, and to trade leisure for income. This conclusion is, perhaps, not surprising. However, a secondary conclusion is that under some plausible restrictions on individual preferences—restrictions that ensure that the substitution effect of greater risk exceeds the income effect—increased tax base risk often changes individual behavior so as to increase the size of the income tax base, while increased tax rate risk always has the opposite effect on the base. It therefore appears that some types of tax uncertainty may actually benefit a government that is only interested in generating more tax revenues. A corollary is that in those cases in which uncertainty increases tax revenues, the government may be able to alter tax rates so that the welfare of its citizens increases, despite the presence of greater uncertainty.

I. Tax Base Uncertainty

A. *Itemized Deductions*

Perhaps the area in which tax base uncertainty is most prevalent is that of itemized deductions. Suppose an individual must choose between two goods X and Y with unit prices, where some random proportion α of expenditures on X is deductible from the income tax base, while expenditures on Y are not. The fixed income I_0 of the individual must equal the sum of expenditures on X, on Y, and on income taxes, which are imposed at rate τ and are equal to $(I_0-\alpha X)$. Utility is assumed to be a function of X and Y. Substitution of $[I_0(1-\tau)-X(1-\alpha\tau)]$ for Y and maximization of expected utility $V\equiv \mathscr{E}U(x,Y)$ gives the following first-order condition for the optimal choice of X^*:

$$(1)\quad \mathscr{E}\{U_X(X,Y)-U_Y(X,Y)\times[1-\alpha\tau]\}=0.$$

Now let there be an "increase in risk" in the random variable α. Michael Rothschild and Joseph Stiglitz (1970) show, when comparing two random variables α and α' with the same expected value, that α' is riskier than α if its distribution can be obtained from that of α by a mean-preserving spread. This definition can be used to determine the impact of increased risk upon the optimal choice of X^*. Suppose that the expression in parentheses in equation (1) is a concave function of α. Then an increase in risk will decrease the optimal choice of X^*, since greater risk will lower the expectation of this expression and, by the second-order condition, the first-order condition can then be restored by a reduction in X^*. Conversely, if the expression is a convex function of α, then an increase in risk will increase X^*.

The impact of tax base uncertainty is in general ambiguous. When α becomes riskier, the individual may substitute away from Y, since uncertainty about α makes consumption of Y a random variable. However, in order to ensure a given amount of consumption of the nondeductible good, she may also increase her expenditures on Y. This ambiguity is illustrated by the expression that determines the convexity or concavity of $\{U_X(X,Y)-U_Y(X,Y)[1-\alpha\tau]\}$. This latter function is convex or concave in α as

$$(2)\quad 2U_{YY}(X,Y)-[1-\alpha\tau]XU_{YYY}(X,Y)+XU_{XYY}(X,Y)\gtrless 0.$$

The first term is negative for a risk-averse individual; the second term is also negative if $U_{YYY}(X,Y)>0$, or if, as is typically assumed, there is nonincreasing absolute risk aversion with respect to the risky good Y, where $A(Y)\equiv -U_{YY}(X,Y)/U_Y(X,Y)$. As long as the third term is less than or equal to zero, inequality (2) is less than zero. When preferences are described by a separable utility function, such as in the Cobb-Douglas or the CES case, $U_{XYY}(X,Y)=0$.

It is therefore likely—though not guaranteed—that the expression in brackets in equation (1) is a concave function of α. Consequently, the optimal choice of X^* declines when α becomes riskier, so that the size of the tax base increases. With no actual change in tax polices, the government is able to increase the size of the expected tax base and the amount of expected tax revenues. Note that it can be shown that tax base uncertainty has identical effects on an individual who must allocate income between wages and fringe benefits, in which wages are fully taxed while some random fraction of employee benefits may also be taxed.

B. *Tax Shelters*

A risk-averse individual is assumed to have a fixed endowment of income I_0 to allocate between sheltered income S and nonsheltered income N. The individual incurs some cost c for each dollar of sheltered income. He must also pay income taxes at rate τ on taxable income, which consists of all nonsheltered income plus some random fraction α of sheltered income. Net-of-tax income I therefore equals $[I_0-\tau(N+\alpha S)-cS]$. The individual chooses N and S to maximize an expected utility function $V\equiv\mathscr{E}U(I)$, subject to the constraint that N and S sum to I_0. The first-order condition from this maximization is

$$\mathscr{E}\{U_I(I)[c-(1-\alpha)\tau]\}=0. \tag{3}$$

Now suppose that α becomes riskier. There are two offsetting effects on the choice of N^*. A risk-averse individual will find sheltered income less attractive because of its greater risk, and respond by substituting toward more N. However, in order to be assured of a given level of income, he must increase sheltered income. The net effect on N^* is therefore uncertain due to conflicting substitution and income effects.

The effect of greater risk on the optimal choice N^* depends upon the concavity or convexity of $\{U_I(I)[c-(1-\alpha)\tau]\}$. If this function is convex, then nonsheltered income will increase; if it is concave, then N^* will decrease. Now this function is convex if

$$-2U_{II}(I)-[(1-\alpha)\tau-c]\times SU_{III}(I)>0, \tag{4}$$

and concave if it is less than zero. In general, the sign of this function is ambiguous. Suppose, however, that the convexity or concavity is examined for extreme values of α. When $\alpha=1$, or all S is included in the base, the sign is positive as long as $U_{III}(I)>0$, which is a necessary condition for nonincreasing absolute risk aversion. When $\alpha=0$, a sufficient condition for the sign to be positive is that relative risk aversion $R(I)\equiv-IU_{II}(I)/U_I(I)$ be nondecreasing and less than one. These restrictions are often made because they imply that the demand for the risky good has an income elasticity that is positive and less than one. They are also satisfied by the Bernoulli and logarithmic utility functions. These restrictions are assumed to apply here, so that increased uncertainty about the income tax treatment of tax shelters will decrease the amount that an individual invests in sheltered income.

C. *Evasion*

The individual is assumed to have a fixed amount of income I_0 that she can allocate between declared income D and evaded, or underreported, income E. Declared income is subject to an income tax at rate τ. Evaded income is not taxed; however, some random fraction α may be detected by the authorities, at which point a penalty π is imposed on each dollar of detected income. The uncertain income I of the individual therefore equals $[D(1-\tau)+E(1-\alpha\pi)]$. Expected utility $V\equiv\mathscr{E}U(I)$ is maximized when declared income is chosen so that

$$\mathscr{E}\{U_I(I)[\tau-\alpha\pi]\}=0. \tag{5}$$

Now suppose that the proportion of evaded income that is discovered becomes more uncertain. Declared income will increase if

$$(6)\quad -2U_{II}(I)-[\tau-\alpha\pi]EU_{III}(I)>0,$$

while D^* will decrease if the inequality is reversed. In general, the sign of this function is ambiguous because the signs of $U_{III}(I)$ and $[\tau-\alpha\pi]$ are both uncertain. The ambiguity reflects the conflicting substitution and income effects that affect D^* when α becomes riskier. Declared income will increase because its return is certain, but evaded income will increase because more must be put into E^* to attain an assured level of income.

However, following an analysis similar to that for tax shelters, it can be shown that sufficient conditions for inequality (6) to be positive are that absolute risk aversion be decreasing and that relative risk aversion be nondecreasing and less than one. If these restrictions apply, then the substitution effect will dominate, and greater risk will increase D^*.

D. *Saving*

The saving decision can be examined using a two-period, life-cycle model. The individual has initial income I_0, which already reflects the income tax liability in that period. He may choose first-period consumption C_1, or may save at an interest rate r in order to consume C_2 in the second period. Tax base uncertainty is introduced by writing C_2 as $(1+r(1-\alpha\tau)(I_0-C_1))$; that is, some random fraction α of interest income is subject to the tax rate τ on interest income. When $\alpha=1$, all interest income is taxable, whereas interest income is excluded from the tax base when $\alpha=0$.

The individual chooses C_1 in order to maximize expected utility, where utility equals $U(C_1)+\rho U(C_2)$ and ρ is the discount factor. The first-order condition that determines the optimal C_1^* is

$$(7)\quad \mathscr{E}\{U_1(C_1)-\rho U_2(C_2)\times[1+r(1-\alpha\tau)]\}=0.$$

Greater uncertainty about α will increase first-period consumption if the term $\{U_1(C_1)-\rho U_2(C_2)[1+r(1-\alpha\tau)]\}$ is convex in α, and decrease C_1^* if it is concave. This depends upon the sign of

$$(8)\quad -2U_{22}(C_2)-C_2U_{222}(C_2)\gtrless 0.$$

In general, inequality (8) is ambiguous in sign: a risk-averse individual will increase C_1 because he dislikes the uncertainty of C_2, whereas he will decrease C_1 in order to be certain of a minimum amount of future consumption. However, a sufficient condition for inequality (8) to be positive is that relative risk aversion $R(C_2)\equiv -C_2U_{22}(C_2)/U_2(C_2)$ be nondecreasing and less than one.

Unlike the earlier cases, increased risk about the base of the interest income tax is likely to reduce the tax base. In this case, expected tax revenues decline because future (taxable) interest income falls.

E. *Labor Supply*

Suppose that the individual must choose the amount of labor L to supply at a normalized wage of unity. Labor income is taxed at a rate τ; however, only a random fraction α of income is included in the income tax base. (Net-of-tax income I is $L(1-\alpha\tau)$.) Tax base uncertainty can be viewed here as riskiness about the proportion of income that will be included in the tax base, whether the income comes from tax shelters, employee benefits, itemized deductions, unreported income, or interest income. Maximization of expected utility $V\equiv\mathscr{E}U(I,L)$ requires that

$$(9)\quad \mathscr{E}\{U_I(I,L)[1-\alpha\tau]+U_L(I,L)\}=0.$$

(Note that $U_L(I,L)<0$.) An increase in uncertainty about α may decrease labor supply because the individual will substitute away from an activity with a riskier return, or greater uncertainty may increase labor supply because she will work more to be sure of reaching some given income. The effect upon

TABLE 1—ALLOCATIVE EFFECTS OF GREATER UNCERTAINTY

	Tax Base Uncertainty	Tax Rate Uncertainty
Itemized Deductions	$\Delta X<0$[a]	$\Delta X>0$[b]
Employee Benefits	$\Delta W>0$[a]	$\Delta W<0$[b]
Tax Shelters	$\Delta N>0$[c]	$\Delta N<0$[c]
Evasion	$\Delta D>0$[c]	$\Delta D<0$[c]
Saving	$\Delta C_1>0$[b]	$\Delta C_1>0$[b]
Labor Supply	$\Delta L<0$[b]	$\Delta L<0$[b]

[a] Sufficient conditions for the inequality are that preferences be separable and that absolute risk aversion be nonincreasing.

[b] Sufficient conditions for the inequality are that preferences be separable and that relative risk aversion be nondecreasing and less than one.

[c] Sufficient conditions for the inequality are that absolute risk aversion be nonincreasing and that relative risk aversion be nondecreasing and less than one.

labor supply depends upon the sign of

$$(10) \qquad 2U_{II}(I,L)+IU_{III}(I,L) + LU_{LII}(I,L) \gtreqless 0.$$

If the inequality is less (greater) than zero, then L^* will decrease (increase).

The sign of inequality (10) is ambiguous. However, if utility is separable between labor supply and income, then a sufficient condition for L^* to decrease is that relative risk aversion of income be nondecreasing and less than one. This condition ensures that the substitution effect exceeds the income effect. It is met if the demand for the risky good has an income elasticity that is less than one.

It seems likely, then, that an increase in tax base uncertainty will decrease labor supply. As with the saving decision, the individual responds to greater risk by decreasing productive activities, and the tax base declines as a result.

The results of tax base uncertainty are summarized in Table 1. An increase in uncertainty clearly affects individual behavior, although the direction of the change is ambiguous because of offsetting income and substitution effects. However, under some plausible—and commonly imposed—restrictions on preferences, the direction of the effects can be determined. An individual often responds to greater tax base uncertainty by choosing more of those activities that increase the size of the tax base. In these cases, greater uncertainty about tax policies induced largely by government itself benefits a government interested in generating revenues. Only in the saving and labor supply decisions does the individual alter behavior so as to decrease the tax base.

II. Tax Rate Uncertainty

The other type of uncertain tax policy is tax rate risk. Here the tax base is assumed to be known with certainty; instead, the tax rate that is imposed on income is risky. Frequent changes in tax rate schedules, the failure to index fully the base and rates of the income tax in a world of unpredictable inflation, proposals for tax reform—all of these factors contribute to tax rate uncertainty. The analysis proceeds by using the same models of individual behavior as in Section I. Now, however, τ is the random variable. Like tax base risk, it is found that tax rate risk alters the various individual decisions. Unlike tax base risk, however, the individual always alters behavior so as to decrease the size of the tax base.

A. *Itemized Deductions*

Greater tax rate risk will increase expenditures on goods that can be deducted if

$$(11) \qquad -2U_{YY}(X,Y)-YU_{YYY}(X,Y) +(I_0-X)U_{XYY}(X,Y)>0,$$

and will decrease X^* otherwise. (The fraction of expenditures on X that is deductible from the base is assumed to equal one.) When preferences are separable, inequality (11) will hold as long as relative risk aversion is nondecreasing and less than one. This restriction on relative risk aversion assures that the substitution effect of greater risk, which reduces expenditures on the now riskier good Y, dominates the income effect.[4] Identical results hold for individual choice between wages and employee benefits.

Unlike tax base uncertainty, tax rate uncertainty is likely to decrease the size of the expected tax base because the individual will substitute toward itemized deductions (or toward employee benefits). Taxpayers will therefore respond to random changes in the rate schedules by paying fewer income taxes. However, removal of the uncertainty will increase expected tax revenues.

B. *Tax Shelters*

Greater τ risk increases or decreases nonsheltered income as

$$(12)\quad 2U_{II}(I)+(c-\tau)NU_{III}(I)\gtrless 0,$$

where $\alpha=0$. The sign of inequality (12) is ambiguous because the signs of both $U_{III}(I)$ and $(c-\tau)$ cannot be determined in general. However, when $c>\tau$, inequality (12) is less than zero if relative risk aversion is nondecreasing and less than one; the sign is also negative when $c<\tau$ if there is nonincreasing absolute risk aversion. Nonsheltered income —or the tax base—is therefore likely to decrease when there is greater tax rate risk because it is nonsheltered income that is fully subject to the random tax rate.[5]

C. *Evasion*

The analysis of tax rate risk under evasion requires that both α and τ be risky. It is assumed that the distributions of α and τ are independent of one another and that only the distribution of the tax rate is subject to a mean-preserving spread.

The effect of increased tax rate risk depends upon the concavity or convexity in τ of the expression in brackets in equation (5). Sufficient conditions for this expression to be concave in τ are that absolute risk aversion be nonincreasing and that relative risk aversion be nondecreasing and less than one. If these conditions hold, then greater tax rate risk will cause the individual to decrease declared income because D becomes riskier than E when τ becomes more random. Tax rate uncertainty is therefore likely to make evasion more attractive to individuals.

D. *Saving*

The effect of greater tax rate risk on saving is a straightforward extension of the earlier analysis because α and τ enter the first-order condition (equation (7)) in an identical way. The conditions under which first-period consumption will increase or decrease are therefore the same for α risk and τ risk; indeed, the distinction between the two types of risk is hazy here. Specifically, C_1^* will increase if relative risk aversion is nondecreasing and less than one. Greater τ risk is therefore likely to increase current consumption and so decrease saving because second-period consumption becomes more uncertain (and less attractive) if the tax rate on interest income is more variable.

E. *Labor Supply*

Tax rate and tax base risk also have the same effect on labor supply. Greater uncertainty about the marginal tax rate makes income riskier, and the individual adjusts by substituting away from income and toward leisure. As long as the income effect of greater risk does not offset this, then labor supply will decline. Conditions sufficient to ensure that L^* declines are that preferences be separable between I and L and that relative risk aversion be nondecreasing and less than one.

---

[4] The assumption that $\alpha=1$ is not completely innocuous. When α does not equal one the convexity or concavity of $\{U_X(X,Y)-U_Y(X,Y)[1-\alpha\tau]\}$ depends upon the sign of $\{-2\alpha U_{YY}(X,Y)-[1-\alpha\tau][I_0-\alpha X]U_{YYY}(X,Y)-[I_0-\alpha X]U_{XYY}(X,Y)\}$. When $\alpha=1$, this expression reduces to inequality $(\neq)$. When $\alpha=0$, however, it reduces to $\{-I_0U_{YYY}(X,Y)-I_0U_{XYY}(X,Y)\}$, so that if there is decreasing absolute risk aversion and separability, then nonsheltered income will increase, not decrease. However, this outcome does not seem very likely. Actual values for α are apt to be closer to one than to zero. Perhaps of more importance, the substitution effect (the $U_{YY}(X,Y)$ term) leads to more X^*, and, as long as the substitution effect is larger than the income effect, X^* will increase with greater τ risk.

[5] These conclusions may be altered for nonzero values of α. See fn. 4.

It should be remembered that tax rate risk in general has ambiguous effects on individual behavior. Still, if the substitution effect of greater tax rate risk is dominant, then government-induced uncertainty about tax rates will have a negative impact on expected tax collections.

III. Welfare Effects of Uncertainty

It is natural to ask how tax uncertainty affects individual welfare. One might suppose that risk-averse individuals will always be made worse off by greater risk, and, if greater base or rate risk is unaccompanied by any other government actions, then it is easy to show that risk does in fact lower welfare. However, the previous sections have demonstrated that uncertainty has allocative effects that alter the tax base. If the government changes the tax rate to maintain constant expected revenues, then there is an additional factor—the change in τ—that affects welfare. Suppose, for example, that greater uncertainty increases expected tax revenues. If the government lowers the tax rate so as to maintain constant expected revenues, then there are two offsetting effects on expected utility: the individual is made worse off by the greater uncertainty, but is made better off by the lower tax rate. It is possible that the latter effect will dominate.

To examine this issue, consider the effect on the welfare of a representative individual of uncertainty about the proportion of deductions that will be allowed. Following Stiglitz (1982), suppose that there are only two equiprobable states of the world. In one the proportion of deductible expenditures is $\alpha+\delta$, whereas in the other the proportion is $\alpha-\delta$, where $\delta>0$.[6] The government chooses δ in order to maximize the expected utility V of the individual, subject to the constraints that the individual behave optimally and that the government collect a constant amount of expected revenues. If at $\delta=0$ it is the case that

$$\left.\frac{dV}{d\delta}\right|_{\delta=0}=0 \quad \text{but} \quad \left.\frac{d^2V}{d\delta^2}\right|_{\delta=0}>0,$$

then the introduction of risk increases individual welfare.

It is straightforward but tedious to show that the first- and second-order conditions for the government optimization, evaluated at $\delta=0$, are

$$\left.\frac{dV}{d\delta}\right|_{\delta=0}=0 \tag{13}$$

$$\left.\frac{d^2V}{d\delta^2}\right|_{\delta=0}=(\tau X)^2U_{YY}(X,Y) -(I_0-\alpha X)U_Y(X,Y)\left.\frac{d^2\tau}{d\delta^2}\right|_T, \tag{14}$$

where $d\tau/d\delta|_T$ is the necessary change in the tax rate caused by δ that keeps expected government revenues constant at T and $d^2\tau/d\delta^2|_T$ is its second derivative. Consequently, a necessary and sufficient condition for greater risk to increase welfare is that

$$\left.\frac{d^2\tau}{d\delta^2}\right|_T<\frac{(\tau X)^2U_{YY}(X,Y)}{(I_0-\alpha X)U_Y(X,Y)}; \tag{15}$$

that is, $d^2\tau/d\delta^2|_T$ must be negative and greater in absolute value that the right-hand side of inequality (15).

Now expected government revenues equal $\tau(I_0-\alpha X)$, so that

$$\left.\frac{d^2\tau}{d\delta^2}\right|_T=\frac{\alpha\tau\dfrac{\partial^2X}{\partial\delta^2}}{(I_0-\alpha X)-\alpha\tau\dfrac{\partial X}{\partial\tau}}, \tag{16}$$

when evaluated at $\delta=0$. If the tax rate is restricted to levels in which a greater tax rate generates more revenues, then the denominator of equation (16) is positive. Consequently, $d^2\tau/d\delta^2|_T$ is negative if and only if $\partial^2X/\partial\delta^2 < 0$. From the first-order condition from the individual's optimization, it can be shown that $\partial^2X/\partial\delta^2 \gtreqless 0$ as

---

[6] Note that the introduction of δ corresponds to a mean-preserving spread in the distribution of α.

$$(17) \quad 2U_{YY}(X,Y)-[1-\alpha\tau]XU_{YYY}(X,Y) + XU_{XYY}(X,Y) \gtreqless 0.$$

Note, however, that this condition is identical to that in inequality (2), which specified the conditions under which greater tax base risk will decrease the optimal amount of itemized deductions; that is, the (necessary) condition under which greater risk will increase welfare is the same as the condition under which risk will increase the tax base. Because it is likely that greater risk will increase the tax base, it is also likely that risk will increase welfare. Although a risk-averse individual dislikes uncertainty, the reduction in the tax rate that the uncertainty permits more than offsets the risk, and welfare increases.

In fact, it can be shown that the necessary condition under which greater risk leads to greater welfare in each of the cases considered is also identical to the condition under which risk generates greater revenues. These conditions suggest that tax base risk will often increase welfare, whereas tax rate risk will always decrease welfare. More generally, the conditions suggest that there are some circumstances under which greater uncertainty about tax policies will have a positive impact on welfare.

IV. Conclusions

In the face of uncertainty individuals will change the ways in which they behave. The effects of tax uncertainty on individual decisions are typically difficult to determine, and represent the net impact of conflicting income and substitution effects. If the substitution effect dominates, it is often the case that greater tax base risk leads to more of those activities that are subject to the income tax; greater tax rate risk always changes behavior so as to reduce the base. Government does not have to make actual changes in its policies to affect the allocation of resources. Instead, actions that simply alter the likelihood of tax policy changes will have repercussions.

Under some circumstances, in fact, uncertain tax policy may be a rational policy for a government that seeks more tax revenues and does not care about the variability of those revenues. In those cases in which greater tax base risk increases the size of the expected tax base, expected tax collections increase with no actual change in the rate or base of the income tax. Because of these base effects, uncertain tax policy may also be a rational policy for a government that seeks greater welfare for its citizens, as noted first by Weiss (1976) and Stiglitz (1982) and more recently by Balcer and Sadka (1982) and Chang and Wildasin (1986). This is a surprising result because the utility of risk-averse individuals declines when uncertainty increases. However, it can occur in those cases in which greater uncertainty increases tax revenues, if the government is allowed to lower tax rates so as to maintain revenues at a constant expected level. If the rate reduction dominates the risk effect, then individuals will actually benefit from greater uncertainty.

It is risky to read too much into these results. Individuals typically make decisions simultaneously, yet these decisions are examined separately here. Similarly, tax base and tax rate risk are examined as separate cases, with base risk absent when rate uncertainty is analyzed, and vice versa. The joint effects of both types of uncertainty may well be different from their individual effects.[7] The results also imply that government may intentionally introduce uncertainty. However, government efforts to introduce risk on a systematic basis may backfire, as taxpayers come to anticipate these actions, and government itself may dislike risk in its budget. Finally, the analysis ignores the potentially large administrative costs from variations in the tax base; tax rate variations are likely to be much less costly.[8]

[7] See, however, the analysis of tax rate risk for evasion.

[8] See Shlomo Yitzhaki (1979) for further discussion.

Still, the results indicate that uncertain tax policies often have a negative effect on expected revenues; tax rate risk in particular always has this effect. At least in these cases, complaints about unpredictable government actions seem well-founded. Not only are individuals made worse off by the randomness, but the government itself loses revenues. More stable policies here would yield a double dividend—on individual welfare and on the government budget.

REFERENCES

Balcer, Yves and Sadka, Efraim, "Horizontal Equity, Income Taxation, and Self-Selection with an Application to Income Tax Credits," *Journal of Public Economics*, December 1982, *19*, 291–309.

Chang, F. R. and Wildasin, D. E., "Randomization of Commodity Taxes: An Expenditure Minimization Approach," *Journal of Public Economics*, December 1986, *31*, 329–45.

Danziger, Sheldon, Haveman, Robert and Plotnick, Robert, "How Income Transfers Affect Work, Savings, and the Income Distribution: A Critical Review," *Journal of Economic Literature*, September 1981, *19*, 975–1028.

Hey, John D., *Uncertainty in Microeconomics*, New York: New York University Press, 1979.

Holmes, Oliver Wendell, Jr., "The Path of the Law," *The Mind and Faith of Justice Holmes*, Max Lerner, ed., New York: Random House, 1943.

Rothschild, Michael and Stiglitz, Joseph E., "Increasing Risk: I. A Definition," *Journal of Economic Theory*, September 1970, *2*, 225–43.

Skinner, Jonathan, "The Welfare Cost of Uncertain Tax Policy," NBER Working Paper No. 1947, June 1986.

Stiglitz, Joseph E., "Utilitarianism and Horizontal Equity," *Journal of Public Economics*, June 1982, *18*, 1–33.

Weiss, Lawrence, "The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax," *Journal of Political Economy*, December 1976, *84*, 1343–52.

Yitzhaki, Shlomo, "A Note on Optimal Taxation and Administrative Costs," *American Economic Review*, June 1979, *69*, 475–80.

Tobin's *q* and the Importance of Focus in Firm Performance

By BIRGER WERNERFELT AND CYNTHIA A. MONTGOMERY*

Using Tobin's q as a measure of performance, we seek to estimate the relative importance of industry, focus, and share effects in determining firm performance. Our methods are analogous to those of Richard Schmalensee and, like him, we find that industry effects account for the majority of the explained variance. However, we also find that firm effects exist in the form of focus effects, that is, narrowly diversified firms do better than widely diversified firms. We interpret this finding as consistent with profit maximization by firms with different factor endowments.

In a thought-provoking paper, Richard Schmalensee (1985) shows that accounting rates of return at the business-unit level are strongly influenced by industry effects, only unimportantly influenced by market-share effects, and not influenced by firm effects. Schmalensee finds the last result somewhat surprising and conjectures (p. 349) that firm effects might exist in the form of focus effects; that is, he suspects that widely diversified (less-focused) firms are unable to transfer their competencies to a host of different markets. The purpose of this paper is to investigate this conjecture.

It is worth noting that characterizing firm effects as focus effects involves a change from looking at individual, random differences to looking at systematic differences across types of firms. In other words, firms may have insignificant individual differences, yet still differ in systematic ways when aggregated according to suitable criteria. A prime candidate for such a criterion is level of firm diversification.

Schmalensee cites managerial literature to the effect that widely diversified firms fail to maximize profits. However, it is not necessarily so. A factor-based theory of diversification (Edith Penrose, 1959; David Teece, 1982) also supports a positive focus effect. According to this theory, diversification is prompted by excess capacity of less-than-perfectly marketable factors. One would expect such factors to return slightly less when used outside the industry first chosen. This does not imply that diversifying firms are not maximizing profits, only that their marginal returns decrease as they diversify farther afield. Like Schmalensee's managerial perspective, this revisionist perspective also predicts a positive focus effect. In contrast, the classical school (which looks at diversification as facilitating collusion) would predict a negative focus effect.

Beyond adding the focus effect, our analysis differs from Schmalensee's in two related ways. Whereas he looked at firms' accounting returns at the industry level, we use Tobin's q as a measure of performance and conduct the analysis at the firm level. In effect we trade more detailed data for a better measure of returns.

Section I develops our research design as it relates to Schmalensee's. Section II outlines our methods. The remainder of the paper describes our results (Section III) and the main implications of the research (Section IV).

I. Research Design

Schmalensee (1985) estimated the model

$$r_{ij} = \mu + \alpha_i + \beta_j + \gamma S_{ij} + \varepsilon_{ij}, \tag{1}$$

where r_{ij} is the accounting rate of return of firm i's operations in industry j, S_{ij} is its market share, the α's are firm effects, the β's are industry effects, μ and γ are constants, and the ε's are disturbances. The results from this model are that industry effects (β's) are important, the market-share effect (γ) is significant but unimportant, and firm effects (α's) add no explanatory power (do not exist).

---

*J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60201. We are grateful to Marvin Lieberman, Mark Satterthwaite, and two anonymous referees for comments, and to Stephen Ross and Trinet for supplying part of the data used in this study. Rob Sartain provided able research assistance.

From this, we would expect that a firm's rate of return r_i can be expressed as

$$(2) \qquad r_i = \Sigma_j \beta_j w_{ij} + \gamma \Sigma_j S_{ij} w_{ij} + \varepsilon_i,$$

where w_{ij} is the percentage of firm i's investments which are made in industry j (since these sum up to one, there is no need for an intercept). If there are firm effects associated with narrowly diversified firms, we expect

$$(3) \quad r_i = \Sigma_j \beta_j w_{ij} + \gamma \Sigma_j S_{ij} w_{ij} + \phi D_i + \varepsilon_i,$$

to perform better than (2), if D_i is a measure of the focus (breadth of diversification) of firm i.

II. Methods

There has recently been considerable criticism of the use of accounting measures of performance (George Benston, 1985). In particular, accounting rates of return are distorted by a failure to consider differences in systematic risk, temporary disequilibrium effects, tax laws, and accounting conventions regarding *R&D* and advertising. These properties are likely to vary more across industries than across firms. That is, if a firm wants to compete in a given industry, it is subject to the risk and disequilibrium effects in that industry and the industry-specific needs for investments in fixed capital, advertising, and *R&D*. Accordingly, the use of such measures will strongly bias an estimation of (1) in favor of industry effects.

As a first step toward avoiding these problems, we chose to use Tobin's q, defined as the capital market value of the firm divided by the replacement value of its assets, as our dependent variable. Theoretically, q is a much more appealing measure than accounting returns. By incorporating a capital market measure of firm rents, q implicitly uses the correct risk-adjusted discount rate, imputes equilibrium returns, and minimizes distortions due to tax laws and accounting conventions.

While q in principle should correct for all the biases noted above, in practice its calculation (Eric Lindenberg and Stephen Ross, 1981) leaves intangible assets out of the denominator, thus overstating the relative performance of firms with large investments in intangibles. The capitalized value of these investments is, of course, very difficult to estimate. However, as a partial correction we include, as independent variables, estimates of a firm's current marketing and *R&D* expenditures, divided by the replacement cost of physical assets (also see Michael Salinger, 1984).

In order to estimate a firm's investment per industry, we multiply its sales by the industry average capital-output ratio. This in turn allows us to construct $\bar{w}_{ij}$, the estimated fraction of firm i's assets that are in industry j.

To measure the extent of firm diversification, we use the concentric index of Richard Caves, Michael Porter, and Michael Spence (1980). The index is given by $D_i = \Sigma_j \bar{w}_{ij} \Sigma_l \bar{w}_{il} d_{jl}$, where d_{jl} is a weight whose value depends on the relations between j and l in the standard industrial classification (SIC) system. The weight is zero if j and l have the same 3-digit SIC code, one if they have a different 3-digit code but identical 2-digit codes, and two if they have different 2-digit codes. This index captures the degree of relatedness between industries, while correlating very strongly with the standard Gort-Herfindahl index (Caves et al., 1980, p. 201).

We estimate the following equations:

$$(4) \quad q_i = \sum_{j=20}^{39} \beta_j \tilde{w}_{ij} + \gamma \sum_{l=1}^{n} S_{il} \bar{w}_{il} + \phi D_i + \varepsilon_i$$

$$(5) \quad q_i = \sum_{j=20}^{39} \beta_j \tilde{w}_{ij} + \gamma \sum_{l=1}^{n} S_{il} \bar{w}_{il} + \phi D_i + \psi_1 \sum_{l=1}^{n} A_l M_{il} / K_i + \psi_2 \sum_{l=1}^{n} R_l M_{il} / K_i + \varepsilon_i,$$

where q_i = an estimate of Tobin's q for firm i;

$\tilde{w}_{ij}$ = the estimated fraction of firm i's assets which are in 2-digit industry j;

S_{il} = firm i's market share in industry l;

$\bar{w}_{il}$ = the estimated fraction of firm i's assets which are in $3\frac{1}{2}$-digit industry l;

D_i = firm i's diversification;

A_l = the ratio of marketing expenditures to sales in industry l;

M_{il} = firm i's sales in $3\frac{1}{2}$-digit industry l;

K_i = the replacement value of firm i's physical assets;

R_l is the ratio of $R\&D$ expenditures to sales in industry l.

To interpret (4), note that the β's are industry effects (estimates of industry q's), γ is the share effect, and ϕ is the focus effect. Following Schmalensee (1985), we can estimate the relative importance of these effects by comparing the R^2's from (4) with regressions having only one or two of the effects. (If we drop the β's, we of course need a constant in the regression.) In (5), we include the partial correction for industry-specific intangible assets. This equation can be subjected to the same analysis as (4); the difference in the resulting estimates of industry effects will give us an idea of the magnitude of industry bias due to accounting practices.

Industry effects are estimated at the 2-digit level. Unfortunately, our data do not allow 3-digit industry estimates. Since average profitability across 2-digit industries is more similar than average profitability across 3-digit or 4-digit industries, this will give a downward bias in our estimates of the importance of industry effects. On the other hand, research on stock prices and industry effects indicates that the bias may be quite small. In his seminal paper, Benjamin King (1966) found that differences between 2-digit industries explained 20 percent of the variance in stock prices on a given day. Later, James Farrell (1974) showed that the same 20 percent could be explained by grouping 2-digit industries into 4 "industry-types." On the other hand, Miles Livingston (1977) explained only 23 percent of the variance using more than 100 different industries. In sum, research on stock prices seems to indicate that estimates of the importance of industry effects are quite insensitive to the level of aggregation. Nevertheless, we should expect that our estimate of the importance of industry will be biased downward. We see no reason, however, why this bias should affect our estimate of the focus effect.

Data sources are detailed in the Data Appendix.

III. Results

Figures 1 and 2 summarize the results of least squares estimation of (4) and (5) and restricted models, excluding one or more of the three effects with which we are concerned. Each arrow corresponds to the imposition of a restriction that one of the three effects discussed above is absent. The number next to each arrow is the probability level (P level) at which an F-test rejects that restriction. (This method of presentation conforms to that of Schmalensee and is used to maximize comparability with his results.)

Note that the adjusted R^2's are quite high. Further, the high P levels generated by tests for share effects (arrows pointing to the left) indicate the absence of share effects in this sample, a result which is particularly strong in the more complete models. In contrast, and consistent with our main hypothesis, tests for focus effects (arrows pointing to the right) show strong significance, especially in the more complete models. As in Schmalensee (1985), the strongest results appear for industry effects which are always significant beyond the .2 percent level.

To analyze the results in another way, Table 1 gives the incremental contributions to the adjusted R^2 of each effect, relative to a model with all three effects. That is, the first cell in Table 1 gives the difference between the adjusted R^2 for the model with industry, focus, and share effects, and the adjusted R^2 for the model with only share and focus effects.

Three results are important. First, like Schmalensee, we find that industry effects account for the majority of the explained variance. In fact, the estimate of 19 percent which emerges from the model without correction for intangible assets, is very close to his and also agrees with the stock price

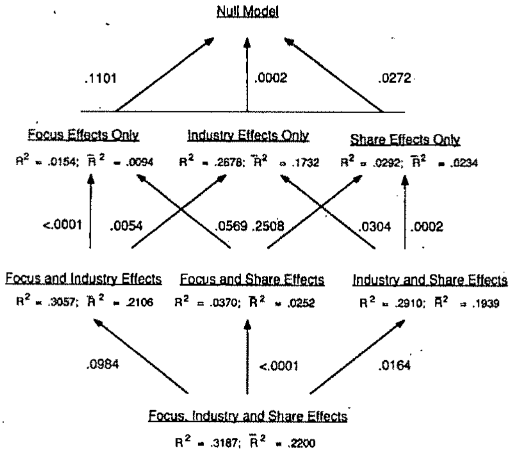

Null Model
.1101
.0002
.0272
Focus Effects Only
$R^2 = .0154$; $\bar{R}^2 = .0094$
Industry Effects Only
$R^2 = .2678$; $\bar{R}^2 = .1732$
Share Effects Only
$R^2 = .0292$; $\bar{R}^2 = .0234$
<.0001
.0054
.0569
.2508
.0304
.0002
Focus and Industry Effects
$R^2 = .3057$; $\bar{R}^2 = .2106$
Focus and Share Effects
$R^2 = .0370$; $\bar{R}^2 = .0252$
Industry and Share Effects
$R^2 = .2910$; $\bar{R}^2 = .1939$
.0984
<.0001
.0164
Focus, Industry and Share Effects
$R^2 = .3187$; $\bar{R}^2 = .2200$

FIGURE 1. MODELS WITHOUT CORRECTION FOR MARKETING AND *R&D* INVESTMENTS

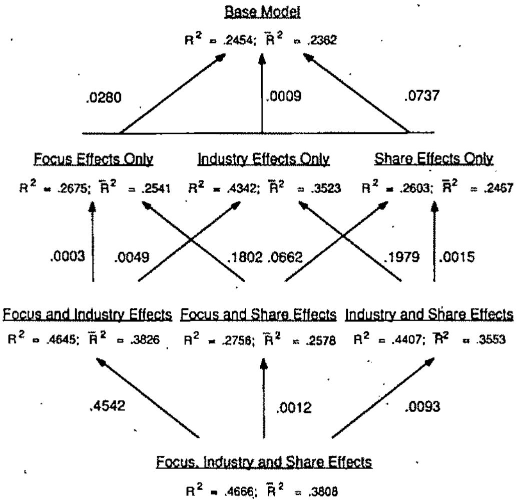

Base Model
$R^2 = .2454$; $\bar{R}^2 = .2362$
.0280
.0009
.0737
Focus Effects Only
$R^2 = .2675$; $\bar{R}^2 = .2541$
Industry Effects Only
$R^2 = .4342$; $\bar{R}^2 = .3523$
Share Effects Only
$R^2 = .2603$; $\bar{R}^2 = .2467$
.0003
.0049
.1802
.0662
.1979
.0015
Focus and Industry Effects
$R^2 = .4645$; $\bar{R}^2 = .3826$
Focus and Share Effects
$R^2 = .2756$; $\bar{R}^2 = .2578$
Industry and Share Effects
$R^2 = .4407$; $\bar{R}^2 = .3553$
.4542
.0012
.0093
Focus, Industry and Share Effects
$R^2 = .4666$; $\bar{R}^2 = .3808$

FIGURE 2. MODELS WITH CORRECTION FOR MARKETING AND *R&D* INVESTMENTS

TABLE 1—ESTIMATED VARIANCE DECOMPOSITION[a]

Effect	Without Correction for Intangible Assets Percentage	With Correction for Intangible Assets Percentage
Industry	19.48	12.30
Focus	2.61	2.65
Share	.94	−.18

[a] Incremental contributions to $\bar{R}^2$, relative to models with all three effects.

research discussed in Section II. The estimate of 12 percent from the more correctly specified model is lower than his, probably because of our coarser industry definitions. Second, focus effects are positive and explain about $2\frac{1}{2}$ percent of the variance. Although Schmalensee did not find pure firm effects, they do appear here in the form of focus effects when firms are represented by their diversification profiles. Third, the importance of industry effects declines by approximately one-third after corrections for industry-related biases.

Concerning the last point, a referee alerted us to the possibility of a real effect of advertising and *R&D* on industry profitability. If such an effect exists and is positive, estimates based on the model (5) are biased against industry effects. This is ultimately a theoretical issue. However, using the depreciation rates of .7 and .9 suggested by Henry Grabowski and Dennis Mueller (1978), the coefficients on advertising and *R&D* should be close to 10/3 and 10. Our estimates are generally below these levels, giving some indication that real effects, if they exist, are small.

The absence of share effects is somewhat surprising, but inspection of Figures 1 and 2 shows that this result is quite robust, despite the multicollinearity between share and focus. Recall also that share effects were unimportant (but significant) in Schmalensee's analysis.

It should be further noted that q by definition is positive such that the distribution of the error term cannot be strictly normal. An examination of the residuals does not, however, reveal problems on this front.

IV. Conclusion and Implications

Using Tobin's q as a measure of performance, we have confirmed Schmalensee's finding that industry effects are the major determinants of firm success. There is some evidence that industry-related measurement bias inflates this estimate. In particular, our results indicate that accounting conventions regarding $R\&D$ and advertising distort estimates of the variation of industry returns. However, after correcting for these biases, the industry effect remains strong.

This brings us to our major result, that firm effects exist in the form of positive focus effects. That is, some differences in performance can be explained by efficiency differences firms experience in transferring competencies to widely varying markets. Interpreted in this way, this finding not only supports the revisionist view, it enriches it, since it also tells us something about the sources of efficiency differences. Further, the fact that we find a positive focus effect contradicts the classical view of diversification as a vehicle for collusion.

DATA APPENDIX

The 1976 values of Tobin's q were generously provided by Eric Lindenberg and Stephen Ross. The method of calculation is described in their (1981) paper in which they publish average values of q for 247 firms in the period 1960–77.

The 1976 estimates of firms' sales and market shares per 4-digit SIC code are from the Trinet/EIS (Economic Information Systems) Establishment Database, made available to us through the Center for the Study of Business Markets at Pennsylvania State University. Trinet/EIS is a subsidiary of Control Data, which provides current sales and market share data on line through Dialog Information Services (file: Trinet Establishment Database).

Data on the replacement value of firms' assets are from 1976 Form 10-K's, filed with the Securities and Exchange Commission. 10-K's are widely available in libraries through the Disclosure Information Group, Bethesda, MD 20816.

Industry data on asset/sales ratios, total marketing expenditures/sales, and company sponsored $R\&D$/sales, at the $3\frac{1}{2}$-digit level, are from the *1976 Line of Business Report*, which is published by the Federal Trade Commission and available to the public.

The dataset itself is available from the authors (Wernerfelt and Montgomery, 1987).

REFERENCES

Benston, George J., "The Validity of Profits-Structure Studies with Particular Reference to the FTC's Line of Business Data," *American Economic Review*, March 1985, *75*, 37–67.

Caves, Richard, Porter, Michael E. and Spence, A. Michael, *Competition in the Open Economy*, Cambridge, MA: Harvard University Press, 1980.

Farrell, James L., Jr., "Analyzing Covariance of Returns to Determine Homogeneous Stock Groupings," *Journal of Business*, April 1974, *47*, 186–207.

Grabowski, Henry G. and Mueller, Dennis C., "Industrial Research and Development, Intangible Capital Stocks and Firm Profit Rates," *Bell Journal of Economics*, Autumn 1978, *9*, 328–43.

King, Benjamin F., "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, January 1966, *39*, 139–90.

Lindenberg, Eric B. and Ross, Stephen A., "Tobin's q Ratio and Industrial Organization," *Journal of Business*, January 1981, *54*, 1–32.

Livingston, Miles, "Industry Movements of Common Stocks," *Journal of Finance*, June 1977, *32*, 861–74.

Penrose, Edith, *The Theory of the Growth of the Firm*, London: Basil Blackwell, 1959.

Salinger, Michael A., "Tobin's q, Unionization, and the Concentration-Profits Relationship," *Rand Journal of Economics*, Summer 1984, *15*, 159–70.

Schmalensee, Richard, "Do Markets Differ Much?," *American Economic Review*, June 1985, *75*, 341–51.

Teece, David J., "An Economic Theory of Multiproduct Firms," *Journal of Economic Behavior and Organization*, March 1982, *3*, 39–63.

Wernerfelt, Birger and Montgomery, Cynthia A., "Tobin's q and the Importance of Focus in Firm Performance: Data," J. L. Kellogg Graduate School of Management, Northwestern University, October 1987.

Federal Trade Commission, *Annual Line of Business Report 1976*, Washington: USGPO.

Securities and Exchange Commission, *1976 Form 10-K*, Washington: USGPO.

Trinet/Economic Information Systems, *Establishment Database 1976*.

Oligopoly and Vertical Integration: Note

By Y. JOSEPH LIN*

There is an extensive literature identifying various incentives for firms to vertically integrate (for example, Kenneth Arrow, 1975; Meyer Burstein, 1960; Fritz Machlup and Martha Taber, 1960; Richard Schmalensee 1973; George Hay, 1973, and others). However, all of these analyses have been conducted for monopolists or competitive firms.[1]

This paper shows that among oligopolists vertical *disintegration* may often be preferred, especially if demands are sufficiently inelastic or if licensing or franchising opportunities are available. Disintegration serves to dampen price competition among manufacturers, thereby leading to higher equilibrium prices and possibly higher profits. The argument is illustrated formally in a simple differentiated-products model, and discussed informally for a homogeneous-product model.

I. The Model

Two manufacturers, x and y, produce a differentiated product. They can choose to distribute the products themselves (vertical integration), or to sell through independent distributors (vertical disintegration). We assume that under disintegration each manufacturer sells through a different, exclusive distributor. For simplicity, assume marginal costs of production and distribution are zero.

A randomly chosen consumer values the two manufacturers' products at v_x and v_y, respectively. We assume $v_x - v_y$ is uniformly distributed on the interval $[-0.5, 0.5]$. The number of consumers is normalized to 1. A consumer either buys one unit of the product from x or y, or does not buy at all. She will buy if $v_x - p_x > 0$ or $v_y - p_y > 0$. For simplicity, we assume that the minimum values of v_x and v_y in the consumer population are large enough to sustain equilibria in which all consumers buy. A consumer will then buy product x if $v_x - v_y > p_x - p_y$ and buy product y otherwise. Assuming that the market shares, or demands, of x and y are both positive, these shares will be $s_x = 0.5 - (p_x - p_y)$ and $s_y = 0.5 + (p_x - p_y)$, respectively.

Four vertical structural configurations are possible: (a) x integrated, y integrated; (b) x disintegrated, y disintegrated; (c) x integrated, y disintegrated; (d) x disintegrated, y integrated.

Any firm facing final consumers (a distributor under disintegration, or a manufacturer under integration) sets retail price p, while a disintegrated manufacturer selling to a downstream firm sets wholesale price w. It is appropriate to view the competition as a noncooperative, two-stage Nash game. To obtain the two-stage Nash equilibrium, the second-stage (subgame) equilibrium manufacturer profits under the four vertical configurations are first found. Then, on the basis of these subgame equilibrium profits, manufacturers' optimal vertical structures are determined in the first-stage game.

Subgame (a). Given p_y, manufacturer x solves

$$\max_{p_x} \left[.5 - (p_x - p_y)\right] p_x.$$

---

*Department of Economics, 2107 CEBA, Louisiana State University and Agricultural and Mechanical College, Baton Rouge, LA 70803-6306. I am deeply indebted to an anonymous referee whose detailed criticisms greatly improved earlier versions of the paper. Suggestions from Francis Lui, Steve Farber, Bob Martin, and John Lunn were of great value. All remaining errors are mine.

[1] A number of authors came close to giving a treatment of endogenous oligopolistic vertical structure. These include Mark McBride (1983), Melvin Greenhut and Hiroshi Ohta (1979). The latter authors, in an interesting footnote, mentioned an endogenous case in which oligopolists ended up integrated.

Given p_x, manufacturer y solves

$$\max_{p_y} \left[.5 + p_x - p_y\right] p_y.$$

The solution to these simultaneous maximization problems gives the equilibrium prices for this subgame: $p_{xa} = p_{ya} = 1/2$. The subgame equilibrium profits are therefore

$$\pi_{xa} = \pi_{ya} = 1/4.$$

Subgame (*b*). Given w_x, w_y, and p_y, distributor x solves

$$\max_{p_x} \left[.5 - (p_x - p_y)\right](p_x - w_x).$$

Given w_x, w_y, and p_x, distributor y solves

$$\max_{p_y} \left[.5 + p_x - p_y\right](p_y - w_y).$$

The solution gives the p's in terms of the w's: $p_x = (1.5 + 2w_x + w_y)/3$; $p_y = (1.5 + 2w_y + w_x)/3$. (Thus, final prices will be higher than those in subgame (a) if the w's are positive, as I will next show.) The market shares in terms of the w's are $s_x = .5 - (w_x - w_y)/3$; $s_y = .5 + (w_x - w_y)/3$. Thus, given w_y, manufacturer x solves

$$\max_{w_x} \left[.5 - (w_x - w_y)/3\right] w_x.$$

Given w_x, manufacturer y solves

$$\max_{w_y} \left[.5 + (w_x - w_y)/3\right] w_y.$$

Solving these simultaneously gives the subgame equilibrium prices: $w_{xb} = w_{yb} = 3/2$; $p_{xb} = p_{yb} = 2$. The subgame equilibrium profits for the manufacturers are therefore

$$\pi_{xb} = \pi_{yb} = 3/4.$$

Subgame (*c*). (Derivations are in fn. 2.)

$$\pi_{xc} = 9/16; \ \pi_{yc} = 3/16.$$

These payoffs are reversed in subgame (d).

TABLE 1—THE FIRST-STAGE GAME BIMATRIX

		x	
		disintegrate	integrate
y	disintegrate	(12,12)	(9,3)
	integrate	(3,9)	(4,4)

The first-stage game is then represented by a bimatrix, whose entries are the pairs of subgame equilibrium profits of manufacturers x and y, respectively, scaled to become integers. The two Nash equilibria of this bimatrix game are (disintegrate, disintegrate) and (integrate, integrate). However, the latter is unstable. Referring to the Table 1 bimatrix, we see that if x first chooses to deviate from (integrate, integrate), then clearly it will be optimal for y to disintegrate as well; x knows this and so will readily disintegrate. Similarly, if y disintegrates first, x will follow suit. Thus, the only "reasonable" Nash equilibrium is (disintegrate, disintegrate).

II. Discussion

The novel idea underlying the preceding results is that vertical disintegration dampens the price competition among manufacturers. Imagine that, under integration, manufacturer x unilaterally increases the retail price p_x by an amount d. This would result in a decrease of x's sale by an amount, say, $\alpha(d)$. ($\alpha(d) = d$ in our simple linear model.) Next, consider the case with disintegration. Imagine that, starting from the same initial situation, the same manufacturer makes the same unilateral increase, d, in its wholesale price w_x. This would lead to an increase of d in the marginal cost of distributor x. If distributor x responded with an equal increase

---

[2] In subgame (c), given p_y, the integrated manufacturer x maximizes $(.5 + p_y - p_x)p_x$ by choosing p_x. Given p_x and w_y, distributor y maximizes $[.5 + p_x - p_y](p_y - w_y)$ by choosing p_y. The FOC's give $p_x = (1.5 + w_y)/3$ and $p_y = (1.5 + 2w_y)/3$. The share of y in terms of w_y is $.5 - w_y/3$. Thus, manufacturer y chooses w_y to maximize $(.5 - w_y/3)w_y$. This implies $w_y = 3/4$ and $\pi_y = 3/16$; $p_x = 3/4$ and $p_y = 1$. These give $\pi_x = 9/16$. For subgame (d), the derivations are exactly similar.

in its retail price p_x, then its sales, and hence those of manufacturer x, would also decrease by $\alpha(d)$. However, distributor x would *not* respond by increasing p_x by d. Distributor x would seek an increase in p_x such that the corresponding increase in *marginal revenue* is also d, so that $MR = MC$ would continue to hold. With a marginal revenue curve steeper than the demand curve, p_x would therefore rise by less than d. Thus, should the vertically disintegrated manufacturer x unilaterally increase its wholesale price by d, it sales would decrease by less than $\alpha(d)$. The same holds for manufacturer y. This means that the degree of responsiveness of sales quantity to manufacturer price changes is lower under disintegration. The effect is similar to making the products of the two manufacturers poorer substitutes for each other.[3] In this way, vertical disintegration dampens the price competition between manufacturers. Notice that the argument does not depend on the linearity of our model and is quite general.[4]

This dampening of the price competition between manufacturers, together with the assumption that total demand for the product is constant in the neighborhood of the equilibrium, results in higher prices as well as greater profits for manufacturers under vertical disintegration in our model. In general, if demand is not sufficiently inelastic, our result may have to be altered. But as long as equilibrium profit of each manufacturer-distribution *pair* under disintegration is greater than that of an integrated manufacturer, oligopolistic manufacturers would still prefer vertical disintegration if there are effective ways to extract downstream profits. Commonly enough, charging lump-sum licensing or franchising fees can effect such extractions. Incidentally, this suggests the existence of a kind of franchising that fosters vertical disintegration among firms.[5]

The dampening-of-competition argument given above applies also when the product is homogeneous. Assume that firms facing final consumers use quantity as strategy, so the downstream equilibrium is Cournot, while a disintegrated manufacturer sets wholesale price w. Then the equilibrium profit of a manufacturer-distributor *pair* under disintegration is *always* larger than that of a manufacturer under integration. The reason is as follows. Under integration, total output is at the Cournot-duopoly level. Under disintegration, total output from the Cournot distributors is smaller, since each will have to incur a marginal cost (wholesale price) $w > 0$. The reason w is not driven to zero (manufacturer's marginal cost) is that each manufacturer's demand is a continuous function of w, despite the product's being homogeneous. The continuity arises because (*i*) a distributor's output is a smooth function of w; and (*ii*) by assumption, a distributor is exclusive and cannot switch to a lower-priced manufacturer. Disintegration thus moves the industry away from the Cournot-duopoly output toward the monopoly output, increasing joint manufacturer-distributor profit. (*Note*: I am very much indebted to an anonymous referee who kindly suggested this very intuitive argument.)[6]

---

[3] In our model, this is seen clearly by comparing the demands of the vertically integrated manufacturers, $0.5 - (p_x - p_y)$ for x and $0.5 + (p_x - p_y)$ for y, with their derived demands in the disintegrated case, $0.5 - (w_x - w_y)/3$ for x and $0.5 + (w_x - w_y)/3$ for y, respectively.

[4] It depends on the marginal revenue curve's being steeper than the demand curve, which will always be true for linear and strictly concave demands. For strictly convex demands, it is easy to show that this is true if the demand is sufficiently *inelastic*.

[5] Franchising has been commonly viewed as a vertical structure intermediate between direct ownership and arm's-length transaction, efficient because grass-roots entrepreneurship is called into play. The kind of franchising introduced here is different. Whenever vertical disintegration generates more upstream plus downstream profit than with vertical integration, upstream firms would first create a *bona fide* vertically disintegrated structure and then use franchise contracts to expropriate downstream profits. Given a disintegrated structure, such franchising only affects the distribution of profit between upstream and downstream firms, but not price and output. It also leads to inefficiency insofar as it fosters disintegration, which tends to raise prices and depress output.

[6] Notice that the increased-profit argument in this paragraph does not apply to the successive *monopoly* model. There, figuratively speaking, one wedge driven between price and marginal cost by an integrated monopolist sends the price to the monopoly level, where industry profit is at a maximum. With vertical disintegration, a second wedge will drive the price to a level

In this homogeneous-product model, one can show that if demand is sufficiently inelastic, then manufacturer profit alone would be greater under disintegration. This can be done conveniently by using the constant elasticity demand: $p = Q^{1/k}$, $k < 0$. (Derivations are straightforward and available upon request.)[7] If a manufacturer can extract a lump sum from its distributor, for example, through franchise fees discussed earlier, there is an incentive for disintegration regardless of demand elasticity.

III. Conclusion

One potentially limiting assumption used in the above analysis must be pointed out—that a distributor cannot switch to a lower-priced manufacturer. This assumption may be valid, for example, if contractual arrangement between distributor and manufacturer is costly, or if valuable mutual knowledge or trust has been developed between the two. In the differentiated product case, necessary product-specific investment by a distributor may make switching even costlier.

Prevailing intuition regarding vertical integration under conditions of fixed-proportions technology is guided by the successive monopoly model: if distribution is not perfectly competitive, a monopolistic manufacturer integrating downstream will earn profit greater than that of a disintegrated monopolistic manufacturer plus that of its independent distributor(s), owing to the elimination of the successive monopoly distortion.

This paper shows that where manufacturing is an *oligopoly* rather than monopoly, it often pays manufacturers to vertically *disintegrate*, basically because selling though independent, imperfectly competitive distributors can dampen the degree of competition between the manufacturers. Such strategic motivations might offer an explanation of franchising that does not rely on various transaction-cost economies.

REFERENCES

Arrow Kenneth, "Vertical Integration and Communication," *Bell Journal of Economics*, Spring 1975, *6*, 173–83.

Burstein, Meyer, "A Theory of Full-Line Forcing," *Northwestern University Law Review*, February 1960, *55*, 62–95.

Greenhut, Melvin and Ohta, Hiroshi, "Vertical Integration of Successive Oligopolists," *American Economic Review*, March 1979, *69*, 137–41.

Hay, George, "An Economic Analysis of Vertical Integration," *Industrial Organization Review*, 1973, *1*, 188–98.

Machlup, Fritz and Taber, Martha, "Bilateral Monopoly, Successive Monopoly, and Vertical Integration," *Economica*, May 1960, *27*, 101–19.

McBride, Mark, "Spatial Competition and Vertical Integration: Cement and Concrete Revisited," *American Economic Review*, December 1983, *73*, 1011–22.

Schmalensee, Richard, "A Note on the Theory of Vertical Integration," *Journal of Political Economy*, April 1973, *81*, 442–49.

---

higher than and *away* from the monopoly level, resulting in lower industry profit.

[7] A Prisoners' Dilemma may arise in this *homogeneous* product model, with each firm preferring to integrate if the other is nonintegrated. The problem can be partially resolved by thinking of the one-period game in our model as a constituent game in either a supergame with a high enough discount rate, or as a stochastically recursive game with a high enough probability of replay. The reason why a Prisoners' Dilemma occurs in some parametrizations of the model is not well understood.

Families and the Economics of Risks to Life

By MAUREEN L. CROPPER AND FRANCES G. SUSSMAN*

To value a project that will reduce risk of death for N persons, one would like to know what each person is willing to pay to reduce his own probability of death, and, in addition, what his friends and loved ones are willing to pay to reduce his risk of death. The second component of willingness to pay (WTP)—the WTP of loved ones—has been investigated by L. Needleman (1976). Own WTP has been examined in a life-cycle context by W. B. Arthur (1981) and by Donald Shepard and Richard Zeckhauser (1982, 1984) under the assumption that individuals have no loved dependents. This note extends their work by examining how the presence of a family alters WTP to reduce one's own risk of death.

In life-cycle models for a single person WTP is the value of the change in expected utility of consumption caused by a change in probability of death. To extend these models to a head of household, utility should be made a function of the head's own consumption and the consumption of his dependents. Utility that the head receives from dependents' consumption while he is alive should increase his WTP since this utility is lost if he dies. A head of household may, however, receive utility from dependents' consumption after he is dead. This portion of lifetime utility—the utility of a bequest—reduces WTP by reducing the disutility associated with dying.

The net effect of dependents on WTP in the life-cycle consumption model thus depends on the strength of the bequest motive. We attempt to infer the strength of the bequest motive by comparing actual survivors' benefits with the estate stream predicted by the life-cycle consumption model. The estimated bequest motive is sufficiently strong that dependents should, on balance, reduce WTP. This conclusion, however, contradicts empirical evidence that the WTP of married persons to reduce risks to their own lives exceeds the WTP of single people (Michael Jones-Lee, 1986). For the life-cycle model to yield this result it must be modified to incorporate the utility received from dependents that is independent of their consumption (Arthur, 1981). This "utility of family existence" term is analogous to a pure "utility of being alive" term in a single person's utility function (Theodore Bergstrom, 1982). Utility that is independent of consumption may justifiably be ignored when the life-cycle model is used to analyze saving and investment behavior, but cannot be ignored when the model is used to make inferences about WTP.

The remainder of this note formalizes the effect of dependents in a life-cycle model in which the head of household can buy actuarially fair life insurance and annuities. The model is solved using an isoelastic utility function and the resulting estate streams are compared with actual survivors' benefits to infer the marginal rate of substitution between consumption and bequest. Given this information, one can compute the WTP of married and single people based solely on utility of consumption. By comparing this difference with empirical estimates of WTP one can infer the value of dependents independent of their consumption. This "family existence value" comprises a large fraction of a head of household's WTP.

I. The Effect of Dependents on Willingness to Pay in the Life-Cycle Consumption Model

We first consider the effect of dependents on WTP in the conventional life-cycle model, ignoring utility that is independent of consumption. To value a change in probability of death at age j we examine the head of household's lifetime expected utility. This consists of two terms—the utility the head receives while alive from his own consumption and that of his family, and the utility he receives from leaving a bequest. Equation (1) gives the head's expected lifetime utility from age k onward,

---

*Associate Professor, Department of Economics, University of Maryland, College Park, MD 20742 (Gilbert F. White Fellow, Resources for the Future, 1986–87), and Economist, U.S. Environmental Protection Agency, Washington, D.C. 20460, respectively. This research was supported by the Computer Science Center, University of Maryland, College Park, and by the U.S. Environmental Protection Agency. We would like to thank Winston Harrington and two anonymous referees for their comments on an earlier draft of this paper.

$$(1)\qquad J_k = \sum_{t=k}^{T} q_{t,k}\alpha^{t-k}m_t u(c_t) + p_{t,k}\alpha^{t-k}V_t(S_t),$$

where $q_{t,k}$ represents the probability of surviving at least to the end of year t, given that one is alive at age k, $p_{t,k}$ is the probability of surviving exactly $t-k$ years, given that one is alive at age k, and α is the inverse of one plus the rate of time preference.

The head's utility of consumption if he is alive at age t is given by $m_t u(c_t)$, where m_t is the number of equivalent adults in the family and c_t is consumption per equivalent adult. $u(\;)$, the period utility function, is assumed increasing, strictly concave, and bounded from below.

The utility of bequest, $V_t(S_t)$, represents the satisfaction that the head receives from his heirs' consumption in the event of his death. This is given formally by

$$(2)\qquad V_t(S_t) = \max_{\{\tilde{c}_{i,t}\}} \sum_{i=t+1}^{T'} \alpha^{i-t}n_{i,t}f_{i,t}u(\tilde{c}_{i,t})$$

$$\text{s.t.} \sum_{i=t+1}^{T'} r^{i-t}\tilde{c}_{i,t}n_{i,t} = S_t, \quad r = 1/(1+\rho),$$

where S_t is the estate in year t, ρ is the rate of return on bonds, and T' is the maximum age to which heirs can live. $n_{i,t}$ denotes the number of equivalent adult heirs in year i, assuming the head dies in year t, and $\tilde{c}_{i,t}$ their per capita consumption. $f_{i,t}$ is the weight that the head attaches to his dependents' consumption after he is dead versus when he is alive.[1]

Following Arthur (1981) and Shepard and Zeckhauser (1982) we assume that the head can purchase both actuarially fair annuities and bonds.[2] By selling $1 of annuities and investing the dollar in bonds, the head can purchase actuarially fair life insurance. To prevent unlimited sale of annuities, consumption expenditure and investment in bonds, s_t, must satisfy (3), where A_k is the stock of annuities held at the end of year k and y_t is earnings at age t,

$$(3)\qquad \sum_{t=k+1}^{T} r^{t-k}q_{t,k}(m_t c_t + s_t) = A_k + \sum_{t=k+1}^{T} r^{t-k}q_{t,k}y_t.$$

We now consider a head of household's willingness to pay at age k for a small change in his conditional probability of death at age j, $\mathrm{WTP}_{j,k}$, and see how this is affected by the strength of the bequest motive. Formally, $\mathrm{WTP}_{j,k}$ is the wealth that can be taken away from the head in year k in response to a change in his conditional probability of dying in year j, $D_j = 1-(q_{j+1,k}/q_{j,k})$, holding utility constant.

In evaluating WTP we label the case in which the head values the consumption of his dependents equally whether he is dead or alive ($f_{i,t}=1$ for all i and t) the strong bequest case. In this case the head purchases life insurance to the point at which each dependent's consumption is the same whether the head lives or dies. Since the utility received from dependents' consumption is independent of D_j, all that the head loses when D_j is increased is the utility of his own consumption, net of the increased cost of consumption, and the present value of lifetime earnings,[3]

---

[1] If dependents receive utility from the head's consumption and the head incorporates this into his utility function, one could argue that $m_t u(c_t)$ should be multiplied by a factor γ; however, this is equivalent to adjusting the weights on the bequest term (the $f_{i,t}$'s), since multiplying J_k by $1/\gamma$ does not alter WTP.

[2] We have also investigated an alternative assumption, namely, that the individual can borrow and lend at the riskless rate of interest but can never be a net borrower. This does not alter our main conclusion, that the WTP of a head of household falls short of a single person's WTP if a family existence term is excluded from the utility function.

$$(4)\quad \mathrm{WTP}_{j,k}=(1-D_j)^{-1}\left[\sum_{t=j+1}^{T} q_{t,k}r^{t-k}\{[u(c_t)/u'(c_t)]-c_t\}+\sum_{t=j+1}^{T} r^{t-k}q_{t,k}y_t\right].$$

Equation (4) implies that a head with a strong bequest motive has a WTP that is unambiguously lower than that of a single person with the same utility function $u(\ )$, earnings stream, and rate of time preference. The single person's WTP is identical in form to (4), but his consumption at each age is higher than the head's, since the head must provide for his family. As long as WTP is increasing in consumption,[4] the head's WTP is lower than that of an otherwise identical single person: dependents unambiguously lower WTP.

For the life-cycle consumption model to predict otherwise, one must either assume a weaker bequest motive, or one must assume that married and single people differ systematically in their preferences. Specifically, married people must, on average, be more risk averse than single people. In the absence of evidence to support the latter, we believe that it is more appealing to view marriage as adding bequest and utility-of-dependents terms to a person's utility function.[5] We therefore maintain the assumption that single and married people do not differ in α or $u(\ )$.

For the life-cycle consumption model to yield the result that dependents raise WTP, one must therefore assume a weaker bequest motive. To simplify matters, suppose that $f_{i,t}=\bar{f}$ for all i and t. If $\bar{f}<1$, the head receives more utility from dependents' consumption when he is alive than after he is dead. This means that the bequest term no longer fully offsets the utility received from dependents' consumption while the head is alive. This adds an extra term to the head's WTP that is not present in the single person's WTP and makes it possible for the head's WTP to exceed the single person's.

To say whether the life-cycle consumption model in reality explains why the WTP of a head of household would exceed the WTP of a single person, we must know the strength of the bequest motive. In Section III we infer $\bar{f}$ by computing estate streams for the model of (1)–(3), under the assumption $u(c)=c^{\beta}$, and comparing them to actual survivors' benefits. Empirical evidence suggests that $\bar{f}$ is high enough to cause the WTP of a head of household to fall short of the WTP of a single person in the conventional model. This suggests that the model must be amended to include utility that is independent of consumption.

II. The Strength of the Bequest Motive and the Importance of Family Existence

In attempting to infer the strength of the bequest motive we have deliberately made assumptions that tend to bias $\bar{f}$ downward, thus strengthening the case for a model in which utility depends solely on consumption. Since the value of $\bar{f}$ consistent with a given estate stream is lower the lower is β, we have chosen a value of β, 0.2, which is at

---

[3] $\mathrm{WTP}_{j,k}$ is obtained by appending (3) to (1) with a Lagrange multiplier and evaluating $-(\partial J_k/\partial D_j)/(\partial J_k/\partial A_k)$. Use of first-order conditions for utility maximization and the fact that $c_t(m_t-1)=\tilde{c}_{i,t}n_{i,t}$ when $f_{i,t}=1$ yields (4).

[4] Since $u(\ )$ is assumed increasing and strictly concave, $(u/u')-c$ is increasing in c, provided utility is positive at the current level of consumption.

[5] The finance literature does not provide conclusive evidence on this point. Richard A. Cohn et al. (1975) using a mail survey of customers of a retail brokerage firm found that the ratio of risky to total assets is higher for single than for married investors. Marshall Blume and Irwin Friend (1975), using the 1962 Survey of Financial Characteristics of Consumers, find that the number of dependents has no effect on the ratio of stocks to net worth.

TABLE 1—COMPARISON OF WILLINGNESSES TO PAY FOR A HEAD OF HOUSEHOLD AND SINGLE PERSON

Age	(1) Estate—OASDI Benefits	(2) Single Person's WTP	(3) Head's WTP, No Existence Term	(4) Family Existence Term
25	21,694	1,823,848	1,496,653	886,075
30	18,312	1,793,385	1,359,648	895,288
35	31,354	1,725,722	1,312,319	867,585
40	50,475	1,619,302	1,223,459	832,437
45	58,269	1,476,938	1,059,643	788,099
50	56,997	1,306,922	877,050	732,606
55	36,863	1,117,229	727,536	665,370
60	11,379	914,550	567,268	577,813
65	−16,945	727,518	420,985	472,866
70	−22,711	614,887	348,349	344,099
75	−16,369	505,553	271,910	187,614

the low end of empirical estimates of this parameter (Sherwin Rosen, 1985). The rate of time preference is assumed equal to the real interest rate, which is 5 percent.

The head of household's earnings at each age are assumed equal to average 1979 earnings (measured in 1981 dollars) of males with 1–3 years of college education (U.S. Bureau of the Census, 1983) multiplied by the probability of employment at each age. The annualized value of these earnings, \$21,627, is approximately equal to median income of full-time male workers in 1979, which is appropriate if one wants to capture an average value of $\bar{f}$. Using mortality rates for white males (U.S. National Center for Health Statistics, 1980), the individual's expected lifetime earnings, discounted to age 18, are approximately \$375,000. Inherited wealth is assumed to be zero.

The head of household is assumed to marry at age 23 and to have children at ages 25 and 28. In computing m_t, each adult 18 and older is assigned a weight of 1.0. Each child receives a weight of 0.3 through age 13 and of 0.62 between 14 and 17 (Walter Dolde, 1978). Children are assumed to leave the household at age 22, and the spouse to die when the head is 80.

To infer $\bar{f}$ the estate streams generated by the model are compared to the present value of actual survivors' benefits, and $\bar{f}$ is varied until the two coincide. For most people, the bulk of survivors' benefits come from Social Security payments (mother's and children's benefits; widow's benefits), real estate, life insurance proceeds, and stocks and bonds. Since data on the last three, by age, are difficult to obtain, we have computed the present value of Social Security survivors' benefits that heirs would receive if the 1979 level of survivors' benefits were to remain constant in real terms.[6] The difference between these and the optimal estate stream when $\beta = 0.2$ and $\bar{f} = 0.75$ appears in column one of Table 1.

If, indeed, $\bar{f} = 0.75$, then the first column of Table 1 should approximately equal the net worth plus life insurance coverage that the average male household head actually possesses. Data on life insurance coverage (American Council of Life Insurance, 1986) suggest that mean coverage per adult male is approximately equal to two years' disposable income. The actual amount of coverage, however, varies signficantly with age, with peak coverage occurring between the ages of 35 and 55 and the amount held declining sharply thereafter. These data are roughly consistent with the first column of Table 1, suggesting the $\bar{f} = 0.75$ if life insurance coverage is the only source of survivors' benefits other than Social Security. To the extent that other sources of benefits exist, $\bar{f}$ exceeds 0.75 and the head's WTP is even lower than that computed below.

---

[6] All computations are based on the 1978 *Social Security Handbook* (U.S. Social Security Administration). They assume that the head's children receive benefits through their 21st year, and that his spouse receives widow's benefits beginning at age 60.

When $\beta = 0.2$ and $\bar{f} = 0.75$ the head's WTP for a change in his current probability of death is below the WTP of an otherwise identical single person at all ages. As indicated in Table 1 (compare columns 2 and 3), this difference is substantial: between the ages of 30 and 70 the head's WTP is less than 75 percent of the single person's WTP. If, as is commonly believed, the WTP of people with dependents exceeds the WTP of people without, this cannot be explained by the conventional life-cycle consumption model.

This model, however, ignores the utility that the head of household receives from his existence and that of his family independently of consumption. Adding such a term to (1) in no way affects the head's consumption and investment decisions, but adds a term to WTP—a term that reflects the existence value of the family—that does not enter the single person's WTP.

To get some idea of the magnitude of the family existence effect, one must know the difference between the willingness to pay of single people and those with dependents for a change in own risk of death. There is, to our knowledge, no published information on this topic. Jones-Lee (1986) reported to us that in a survey of willingness to pay to reduce risk of dying in a traffic accident, he and his colleagues found the willingness to pay of married people to be 7 to 28 percent higher than the WTP of single people for the same risk reduction.[7] Adding a term $g(M_t) = 2.25M_t$ to the period utility function, where M_t denotes the number of dependents in year t, produces the family existence value terms in column 4 of the table, which are consistent with the Jones-Lee et al. findings.[8] When these terms are added to the other components of the head's willingness to pay, the head's willingness to pay exceeds the single person's by approximately 25 percent between the ages of 25 and 65. The family existence term is substantial: it constitutes about 40 percent of total WTP between ages 25 and 40 and 53 percent of WTP at age 60.

III. Summary and Conclusion

As Bergstrom (1982) has pointed out, an individual who has preferences over lotteries involving lifetime consumption streams and length of life will, under certain simplifying assumptions, have a utility function equal to the discounted sum of expected utility of consumption each period, plus a term that is a function of survival probabilities alone. The latter term, however, has been ignored in life-cycle WTP models, presumably because it is difficult to parameterize. This note suggests that this term should not be ignored, especially when the life-cycle model is modified to incorporate the effect of dependents on a person's WTP to reduce his own risk of death.

If one does ignore the term and computes WTP using a standard life-cycle model with bequest, the WTP of a married person falls short of the WTP of an otherwise identical single person. Since empirical results suggest that the WTP of married people exceeds the WTP of single people, one must either assume that a married person receives utility from the existence of his family which a single person does not, or one must assume that his degree of risk aversion is sufficiently higher than that of the single person. In the absence of conclusive evidence regarding the latter, our results suggest that a "utility of existence" term should be added to the life-cycle consumption model if the latter is to be used to analyze WTP.

REFERENCES

Arthur, W. B., "The Economics of Risks to Life," *American Economic Review*, March 1981, *71*, 54–64.

Bergstrom, Theodore C., "When Is a Man's Life Worth More Than His Human Capital?," in M. W. Jones-Lee, ed., *The*

---

[7] These percentages represent differences in mean responses to questions 18 and 20 of the survey, which is described in Jones-Lee, M. Hammerton, and P. Philips (1985).

[8] We assume that the head of household and single person receive the same utility from their own existence, hence only the utility received from dependents affects the difference between the two willingnesses to pay.

Value of Life and Safety, Amsterdam: North-Holland, 1982.

Blume, Marshall E. and Friend, Irwin, "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *Journal of Finance*, May 1975, *30*, 585–603.

Cohn, Richard A., Lewellen, Wilbur G., Lease, Ronald C. and Schlarbaum, Gary C., "Individual Investor Risk Aversion and Investment Portfolio Composition," *Journal of Finance*, May 1975, *30*, 605–20.

Dolde, Walter, "Capital Markets and the Short Run Behavior of Life Cycle Savers," *Journal of Finance*, May 1978, *33*, 413–28.

Jones-Lee, M. W., Hammerton, M. and Philips, P. R., "The Value of Safety: Results of a National Sample Survey," *Economic Journal*, March 1985, *95*, 49–72.

______, Personal communication, November 1986.

Needleman, L., "Valuing Other People's Lives," *The Manchester School of Economics and Social Studies*, December 1976, *44*, 309–42.

Rosen, Sherwin, "The Value of Longevity and Life Expectancy," mimeo., Chicago: University of Chicago, 1985.

Shepard, Donald S. and Zeckhauser, Richard J., "Life Cycle Consumption and Willingness to Pay for Increased Survival," in M. W. Jones-Lee, ed., *The Value of Life and Safety*, Amsterdam: North-Holland, 1982.

______, "Survival versus Consumption," *Management Science*, April 1984, *30*, 423–39.

American Council of Life Insurance, *Life Insurance Fact Book*, Washington: American Council of Life Insurance, 1986.

U.S. Bureau of the Census, "Lifetime Earnings Estimates for Men and Women in the United States: 1979," *Current Population Reports*, Series P-60, No. 139, Washington: USGPO, 1983.

U.S. National Center for Health Statistics, *Vital Statistics of the United States—1978*, Washington: USGPO, 1980.

U.S. Social Security Administration, *Social Security Handbook*, 6th ed., Washington: USGPO, 1978.

Some Welfare Implications of Job Mobility in General Equilibrium

By EARL L. GRINOLS AND STEVEN J. MATUSZ*

A reading of public opinion indicates that intense (and presumably unfair) foreign competition contributes to higher rates of domestic unemployment for certain sectors of the economy. Furthermore, it is a common belief that the jobs which are lost due to foreign competition are the "best" jobs. Even though a growing body of literature exists, which integrates various forms of unemployment into the traditional models, little has been done to address the issue that foreign trade affects worker welfare by causing workers to take other types of work.[1] The purpose of this note is to present a framework in which different "jobs" and movement between them can be considered within the framework of a simple general-equilibrium model. In this context, we verify that it is possible for trade liberalization to cause a worker to move into an inferior job. However, with workers moving between job types the implications of a policy change are not monotonic with the size of the change. We show that if liberalization is a stepwise procedure, the initial steps might reduce worker welfare up to a point, while further liberalization will then lead to a higher level of utility than earned by the worker in the pre-liberalization state. After trade has been completely liberalized, the worker may be better off or worse off as compared with the initial situation.

A significant finding of our model for policy purposes is that the response of the economy to interventions that influence relative product prices (for example, excise taxes, restrictions on foreign trade) is not fully predictable when only data that are observable immediately prior to the change are taken into consideration. Unobservable factors relating to the worker's mobility between jobs now plays a role. The result is that the magnitude of the policy intervention can be just as important as the direction of the intervention for the determination of the impact on the economic welfare of any given agent.

I. The Model

We assume that there are two sectors of production (X and Y). Output is produced by the completion of a number of different tasks or "jobs", according to a specified technology. Workers can perform all or some of the jobs with varying degrees of efficiency.[2] For simplicity, we assume that there are only two distinct jobs, and that both jobs are required in fixed proportions for the production of each of the goods.

---

*Department of Economics, University of Illinois, Urbana, IL 61801, and Department of Economics, Michigan State University, East Lansing, MI 48824, respectively. We would like to thank the anonymous referee, the participants of seminars at Michigan State University, the University of Illinois, and the University of Wisconsin for their comments on an earlier version of the paper. A special thanks is due to Daniel Hamermesh for directing our attention to various references in the literature. Responsibility for any remaining errors, of course, is our own.

[1] See, for example, Gary Fields and Earl Grinols, 1987; John Harris and Michael Todaro, 1979; or Richard Brecher, 1974, for the incorporation of exogenously imposed minimum wages. Steven Matusz, 1985, analyzes a two-sector model with implicit contracts. See also Lanny Arvan and Francoise Schoumaker, 1987, for an application of the efficiency-wage model to international trade.

[2] The idea of assigning workers to jobs based upon their comparative advantage has a long and distinguished history in the field of labor economics. For example, Jan Tinbergen, 1947; A. Roy, 1951; Hendrik Houthakker, 1974; and Sherwin Rosen, 1978, 1982, all use this paradigm. In addition, Michael Waldman, 1974, has a model where workers are assigned to jobs, though there exists no comparative advantage in his model. In all instances, the primary focus is on the observed distribution of earnings relative to the distribution of labor skills.

Since both sectors are assumed to be perfectly competitive, the wages paid for each job will adjust so that profit is nonpositive in equilibrium. If both goods are produced in positive quantities, profit must be zero:

$$P = a_{x1}W_1 + a_{x2}W_2, \tag{1a}$$

$$1 = a_{y1}W_1 + a_{y2}W_2. \tag{1b}$$

In equations (1a) and (1b), the a_{ij} represent the number of efficiency units of job j required to produce one unit of good i. The wage paid for job j (per efficiency unit) is represent by W_j. Good Y is the numeraire, which we indicate by setting its price equal to one.

There can be any number of factors of production capable of performing either or both jobs. A "factor of production" in this context can be an individual worker or unit of some physical input such as a machine. For exposition, we will focus the analysis upon a typical worker. We normalize this worker's ability by assuming that he can supply exactly one efficiency unit of service to job 1, but only e units of service to job 2, where $e \in (0,1)$.

When faced with wages W_1 and W_2, the worker will choose to work at job 1 (job 2), if and only if $W_1 > eW_2$ ($W_1 < eW_2$). Only if $W_1 = eW_2$ will the worker be indifferent with respect to which job he holds.

We end our description of employment by noting that the labor market in this model is distinguished from a model with homogeneous labor because there are in principle as many factors of production as there are workers with different ability coefficients. In consequence, the work force is heterogenous, even though in equilibrium workers of different types may choose to work in common job descriptions. Because there are only two job descriptions in the present framework, this must happen if there are three or more factors with distinct values for e.

II. Wage and Price Movements

Since we have assumed fixed-coefficients technology, it is possible to solve equations (1a) and (1b) to find W_1 and W_2 as linear functions of P. Upon doing so, we find that

$$W_1 = \frac{a_{y2}P - a_{x2}}{a_{x1}a_{y2} - a_{y1}a_{x2}}, \tag{2a}$$

$$W_2 = \frac{a_{x1} - a_{y1}P}{a_{x1}a_{y2} - a_{y1}a_{x2}}. \tag{2b}$$

Without loss of generality, assume that it is sector X that uses job 2 relatively intensively. The denominators of (2a) and (2b) are therefore negative. As such, W_1 varies inversely as the relative price of X increases, while, W_2 varies directly.[3]

By using (2a) and (2b), we can solve for

$$P^* = ea_{x1} + a_{x2}/ea_{y1} + a_{y2}, \tag{3}$$

the relative price at which the typical worker would be indifferent between jobs. For values of $P < P^*$, the worker will offer his labor services in job 1. As the relative price of X increases, the wage offered for job 2 increases as well, while that offered for job 1 falls. For $P > P^*$, it is in the worker's interest to seek employment in job 2.

III. Trade Liberalization and Real Income

We now investigate the implications of the model for job changes and the welfare of labor.

A. *Welfare*

To do this in a concrete fashion, let us postulate that preferences are represented by a Cobb-Douglas utility function, where α is the share of income spent on good X. From this assumption, we can write the indirect utility function as

$$V(P, I) = AP^{-\alpha}I, \tag{4}$$

---

[3] The domestic relative price of good X lies between a_{x1}/a_{y1} and a_{x2}/a_{y2} if both goods are produced. This follows from (2a), (2b), and the nonnegativity restriction on the wages.

where I represents income and A is a constant term. Furthermore, assume that the country under consideration is a small importer of X, so that it cannot influence world prices by its own actions.

Assuming that $P < P^*$, we know that the individual will be employed in job 1 under free trade.[4] Define *ad valorem* tariff t^* as the tariff for which the individual will be indifferent between jobs. Solving the following equality for t^*,

$$(5) \qquad P(1+t^*) = P^*,$$

gives the relation

$$t^* = \frac{e(a_{x1} - a_{y1}) + (a_{x2} - a_{y2})}{P(ea_{y1} + a_{y2})}.$$

For tariffs larger than t^*, the individual will be employed in job 2.

We can also define tariff t^{**} as the tariff for which the individual earns the same real income with protection as he would earn if free trade prevailed. Such a tariff is larger than t^*. This follows because the increase in the relative (domestic) price of good X that is due to the imposition of the tariff reduces the (real) wage paid for job 1 while increasing the (real) wage paid for job 2. At t^*, the individual's real income has fallen to its lowest level and he changes jobs. Further increases in the tariff act to raise the individual's real income, since he is now employed in job 2.

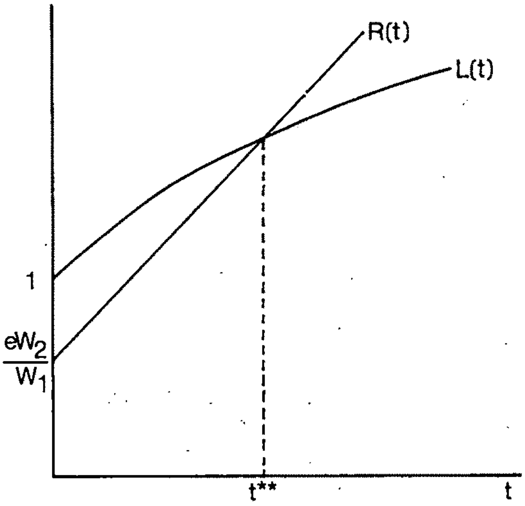


FIGURE 1

To solve for t^{**}, set $V(P,I)$ equal to $V(P(1+t^{**}), I^{**})$, where $I = W_1$ and $I^{**} = eW_2$. From (2) and (4), we obtain the transcendental equation,

$$(6) \quad (1+t^{**})^{\alpha} = \frac{e(a_{x1} - a_{y1}P)}{a_{y2}P - a_{x2}} - \left\{ \frac{ea_{y1}P}{a_{y2}P - a_{x2}} \right\} t^{**}.$$

When evaluated at $t = 0$, the left-hand side of this equality reduces to 1, while the right-hand side becomes eW_2/W_1. This latter term is less than 1 by assumption. Similarly, for large values of t^{**}, the right-hand side exceeds the left. Given this information, it is easy to illustrate the solution to equation (6). We do so in Figure 1, where $L(t)$ and $R(t)$ represent the left- and right-hand sides of equation (6) as functions of t

For $t \in (0, t^{**})$ the individual receives a lower level of real income than under free trade. For $t = t^{**}$, the individual receives the same level of real income with limited trade as with free trade. Finally, if $t > t^{**}$, the individual's real income is higher with protection.[5]

---

[4] From (3), P^* is a function of e and therefore differs from worker to worker. If both goods are to be produced, we know that there must exist at least some workers for whom $P \leq P^*$ and some workers for whom $P \geq P^*$. This follows since both types of jobs are required to produce both goods. The ensuing analysis pertains to those workers for whom $P < P^*$, and where P^* is smaller than the autarkic price. These are the only workers who might switch jobs as the level of protection varies.

[5] If tariff revenue is distributed back to the worker, his welfare will also depend on the amount of revenue

B. *The "Good" Job and the "Bad" Job Refined*

Suppose now that there is initially a moderate level of protection $\bar{t}$ in place, where $t^* < \bar{t} < t^{**}$. In this situation, the individual is employed in job 2 (since $\bar{t} > t^*$), but earns a lower level of real income than he would if there were no tariff (since $\bar{t} < t^{**}$). Furthermore, since $eW_2 > W_1$ we know that $W_2 > W_1$. This defines job 2 as the higher-paying (or "better") job.

C. *Policy Effects on Job Type and Welfare*

Suppose now that the tariff is reduced to t^*. This movement makes the individual worse off, since real income is minimized at t^*. For all tariffs smaller than t^*, the individual is employed in job 1 and marginal reductions in the tariff will increase the welfare of the individual, as the tariff reduction increases the real wage paid for job 1. This means that the person may or may not be better off than under the initial situation. The comparison to determine whether improvement has occurred is with respect to the pre-liberalization position where $t^* < \bar{t} < t^{**}$. This can be shown with respect to Figure 2. Ultimately, greater liberalization will lead to a higher real wage for the individual than he earned initially. This follows since the real income associated with free trade is higher than that associated with the initial level of protection. Thus, in general, the real wage as measured by the value of the indirect utility function varies nonmonotonically with the tariff as shown in Figure 2.

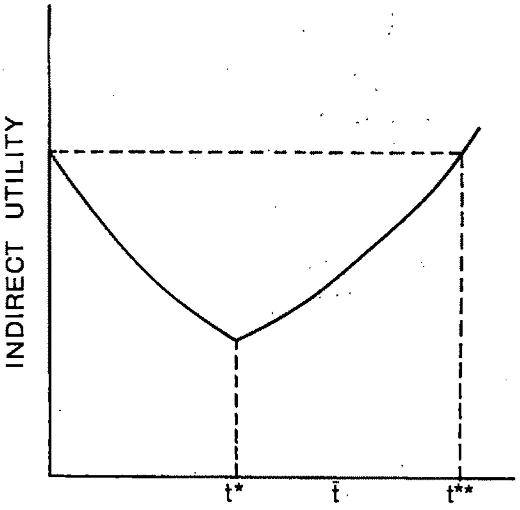


FIGURE 2

IV. Discussion

From Figure 2, we see that the utility level of the typical individual achieves a global maximum at free trade, if the prohibitive tariff is less than t^{**}. If the prohibitive tariff is larger than t^{**}, the global maximum is attained at autarky. Were our model to have had more than two jobs, it is possible to show that there could be another turning point in the relationship between the tariff level and real income. As such, there could be a strictly positive tariff that is not prohibitive which achieves a global maximum for the individual. The analysis culminating in Figure 2 clearly suggests that the welfare consequences of a particular policy depend upon both the direction and magnitude of that policy. For any initial tariff between zero and t^{**}, the individual can benefit by a sufficient liberalization, or by a sufficient increase in protection. Furthermore, the direction of welfare movement for a small change in protection is not a good indicator about the ultimate impact of liberalization.

---

he receives. Depending on the size of his share it is conceivable that he could be better off with any level of protection as compared with free trade. This would happen if the changes in his real income from tariff revenue outweigh the changes in his real income from wages. Suppose that the individual under consideration receives a share of the tariff revenue ϕ, where $0 \leq \phi \leq 1$. In this case, his *real* income is $AP^{-\alpha}(1+t)^{-\alpha}(W_1+\phi tPM)$ if $t \leq t^*$, and $AP^{-\alpha}(1+t)^{-\alpha}(eW_2+\phi tPM)$ if $t \geq t^*$, where M represents the quantity of imports. Since real income now depends on all of the terms involved, an increase in the tariff rate when $t < t^*$ may or may not increase tariff revenue sufficiently to offset the fall in the individual worker's real wage. The precise situation that applies will depend upon how tariff revenues and ϕ vary with the tariff rate.

The manner in which an individual in this model views a particular policy cannot be predicted based upon either sector of employment or type of job held. The relevant determinant is the worker's adaptability to changing conditions, as summarized by the unobservable parameter e. In particular, we see from Figure 1 by raising the slope and intercept of $R(t)$ that increasing the worker's ability to perform job 2 (i.e., increasing e) will reduce t^{**}, the tariff at which the worker is indifferent between protection and free trade. In the case considered here, workers will be for or against a liberalization based on their flexibility to assume other types of jobs (high e or low e). Furthermore, we can infer from equation (3) and the assumption that sector X is relatively intensive in the use of job 2 that increasing the worker's ability to do job 2 increases the person's welfare for every tariff level. This is quite intuitive. With greater ability to perform job 2, less inducement is required before a worker will shift out of a sinking-wage job. In terms of Figure 2, this means that both t^* and t^{**} shift to the left, and real income is now higher (or no lower) for every tariff in excess of the "new" t^* as compared with the previous situation. The relationship between the high e worker's welfare and the low e worker is given in Figure 3.

There are clearly an infinite number of possibilities such that trade liberalization would cause the worker with low ability in job 2 to switch to job 1 and increase his real income in the process, while the high-ability worker would continue to hold job 2, even though the real wage for that job has fallen. A factor may stick with its current line of work, even when the wage is falling, if no better alternative is available. At a certain point, however, the factor will abandon a sinking-wage job and attach itself to different employment for which the wage is rising with the policy. A policy change which was harmful to the factor at low levels may therefore be helpful at higher levels, and vice versa, depending on the mobility of the factor as determined by e.

We began this note by pointing to the notion that there might be good jobs and bad jobs. From an economist's point of view, a good job would be a high-paying job. We have shown that trade liberalization can cause workers to move out of high-paying jobs and into low-paying jobs, as long as their relative ability in the high-paying job is low enough.[6] This is an intuitively appealing result. As the demand for a particular skill wanes as a result of intensified foreign competition or trade liberalization, two things happen. First, the wage paid for that skill falls relative to other wages. Second, the workers of marginal ability in that skill or with relatively greater skills in other types of work move into other jobs.

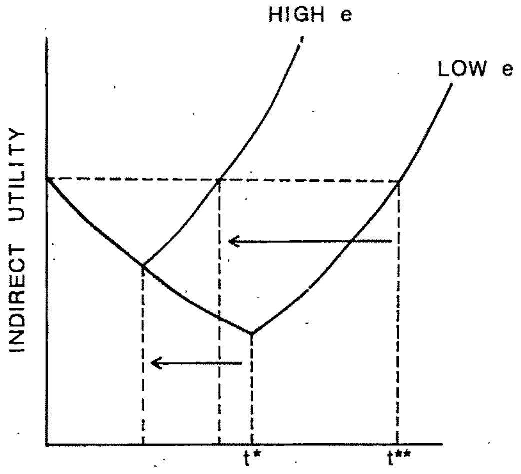


FIGURE 3

Modeling occupational mobility in this way permits us to capture a widely perceived phenomenon in fairly simple terms. Exploration of all of the implications of such mobility are beyond the scope of this note, but it is fairly easy to see that an enlarged model of occupational mobility will alter the general-equilibrium characteristics of the economy, in accordance with the abilities of the labor force over a range of alternative jobs. This has implications for education which is presumably the process that endows workers with high values of e over a range of jobs.

---

[6] While $W_1 > eW_2$, it is still possible that $W_1 < W_2$. That is, job 2 continues to be the high-paying job for the inframarginal workers.

What began as a study of job loss, therefore, becomes an argument for flexibility, and broad-based skills as a hedge against loss in welfare from economy shocks.

REFERENCES

Arvan, Lanny and Schoumaker, Francoise, "Is Public Opinion Right and Conventional Economics Wisdom Wrong About Commercial Policy?," unpublished manuscript, University of Illinois, 1987.

Brecher, Richard, "Minimum Wage Rates and the Pure Theory of International Trade," *Quarterly Journal of Economics*, February 1974, *88*, 98–116.

Fields, Gary and Grinols, Earl, "Import Competition in the High-Wage Sector and Trade Policy Effects on Labor," unpublished manuscript, University of Illinois, 1987.

Harris, John and Todaro, Michael, "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, March 1970, *60*, 122–42.

Houthakker, Hendrik, "The Size Distribution of Labour Incomes Derived from the Distribution of Aptitudes," in W. Sellekaerts, ed., *Econometrics and Economic Theory: Essays in Honour of Jan Tinbergen*, White Plains, NY: International Arts and Sciences Press, 1974, 177–187.

Matusz, Steven, "The Heckscher-Ohlin-Samuelson Model with Implicit Contracts," *Quarterly Journal of Economics*, November 1985, *100*, 1313–29.

Rosen, Sherwin, "Substitution and Division of Labor," *Economica*, August 1978, *45*, 235–50.

______, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, Autumn 1982, *13*, 311–23.

Roy, A., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, June 1951, *3*, 135–46.

Satinger, Michael, "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica*, May 1975, *43*, 455–68.

Tinbergen, Jan, "Some Remarks of the Distribution of Labour Incomes," (original in Dutch, 1947), translation in *International Economic Papers*, *1*, 1951, 195–207.

Waldman, Michael, "Worker Allocation, Hierarchies and the Wage Distribution," *Review of Economic Studies*, January 1984, *51*, 95–109.

Is Increased Price Flexibility Stabilizing?: Comment

By STEPHEN R. KING*

It is widely believed that output in the United States has been more stable since World War II than it was before World War I. Unless this belief results from the use of inaccurate data (Christina Romer, 1986), then macroeconomic theory should help us to understand the phenomenon. J. Bradford De Long and Lawrence Summers (1986a, b) have argued that decreased flexibility of prices explains increased output stability. In their analysis, a positive shock to aggregate demand increases the expected rate of inflation, which reinforces the initial shock by reducing the expected real rate of interest. The stickier prices are, they argue, the smaller the destabilizing inflation effect, and the more variable output is. Their analysis assumes that current prices do not respond to the demand shock. If the current price level adjusts, then there is an additional effect to consider. The increase in the price level decreases real balances, and hence reduces the output movement by increasing the interest rate. It is this effect that presumably gave rise to the "conventional wisdom" that price flexibility attenuates output variation.[1]

In this note, I reassess the contribution of sticky prices to output stability in two main sections. In Section I, I suggest a contracting model that contains De Long and Summers' model as a special case. In Section II, I, like De Long and Summers, use numerical simulations to determine whether price flexibility is stabilizing.

The results are unequivocal. A natural definition of increased flexibility is an increase in the proportion of workers whose wages are determined on spot markets rather than contract ones. An increase in such flexibility is unambiguously *stabilizing* in response to shocks to either aggregate demand or supply. The De Long and Summers' definition of increased flexibility is an increased responsiveness of wages to expected output. We show that an increase in flexibility of this variety is unambiguously stabilizing in response to supply shocks, and will also be stabilizing in response to demand shocks unless the proportion of wages and prices determined under contract is very high. What is more, both types of price flexibility are stabilizing if the monetary authority targets either actual or expected nominal GNP. De Long and Summers' finding that price flexibility can be destabilizing stems from their assumption that there is no contemporaneous response of any prices or wages to demand shocks.[2]

In the absence of further information about either the precise nature of changes in price flexibility or the nature of policy responses, it would be premature to agree with De Long and Summers' claim that their conclusion is robust. Indeed, if wages or prices are allowed to respond at all to output fluctuations in the current year, or if monetary policy attempts to stabilize nominal output, then the evidence suggests that increased price flexibility is more likely to be stabilizing than destabilizing.

I. Model Specification

We make two modifications to the Taylor (1980) contracting model. First, like De Long and Summers, we assume that aggregate demand depends negatively on the anticipated real interest rate. On the supply side, we assume that a fraction of the labor force set their wages on the spot market, while the remainder have two-period overlapping contracts.

---

*Financial Studies Department, Federal Reserve Bank of New York, New York, NY 10045. I am grateful to J. Bradford De Long and Robert J. Gordon, for comments. The views expressed herein are those of the author and are not the official views of the Federal Reserve Bank of New York or the Federal Reserve System.

[1] Recently, for example, in John B. Taylor (1986a, b).

[2] This point has also been made independently in a slightly different model by Steve Ambler and Louis Phaneuf (1987).

The IS curve for the model is simply

$$(1)\quad y_t = by_{t-1} - A(i_t - ({}_tp_{t+1} - p_t)) + u_t^d,$$

where

$$u_t^d = \mu u_{t-1}^d + z_t,$$

where y, p, and i denote the log of the deviation of output from potential, the log of the price level, and the nominal interest rate, respectively. The notation ${}_{t-i}X_{t+j}$ denotes the expectation of X_{t+j} conditional on information at time $t-i$, and z_t is a white noise process.

The LM curve is derived from a conventional log-linear money demand function[3]

$$(2)\quad m_t^d = p_t + y_t - \sigma i_t + v_t^d.$$

Equations (1) and (2) can be combined with an assumed value of the money supply to yield an aggregate demand curve:[4]

$$(3)\quad y_t = 1/(1+A/\sigma) \times \{by_{t-1} - A((1+1/\sigma)p_t - {}_tp_{t+1}) + A/\sigma(m_t^s - v_t^d) + u_t^d\}.$$

We examine the effects of two classes of monetary rule. One assumes that the supply of money is a linear function of the nominal interest rate,[5]

$$m_t^s = \beta i_t.$$

We also examine the case in which monetary policy is directed toward stabilizing either actual or expected nominal GNP.

On the supply side, a proportion θ of wages are determined with two-period contracts similar in spirit to those in the Taylor (1980) model,

$$(4)\quad w_t^c = 1/2({}_{t-1}w_t + {}_{t-1}w_{t+1}) + g/2({}_{t-1}y_t + {}_{t-1}y_{t+1}) + u_t^s,$$

where w_t is the weighted average of spot and preexisting contract wages in period t. The remaining fraction $(1-\theta)$ of the labor force determine their wages in real terms on a spot market, where the wage is determined as[6]

$$(5)\quad w_t^s = gy_t + p_t.$$

The price level is then determined as a markup over a weighted average of spot wages, and current and lagged contract wages, as

$$(6)\quad p_t = (1-\theta)w_t^s + \theta/2(w_{t-1}^c + w_t^c).$$

The model is then composed of the supply side given by equations (4)–(6), the aggregate demand curve (3), a money supply function, and the restrictions implied by rational expectations. De Long and Summers' model is seen to be a special case of this model, in which b is zero, and θ is unity—so that wages and prices are independent of unanticipated output in the current year.

At this stage, we can define two different types of flexibility. Increased flexibility can naturally be modeled as a decrease in θ, the proportion of wages determined under contract. Increased flexibility of this type is likely to be stabilizing since the lower θ is, the larger will be the (attenuating) contempora-

---

[3] The income elasticity of money demand is assumed to be unity to reduce the number of parameters to be varied in the simulations of Section II.

[4] When b is zero, this is exactly equivalent to the formulation in De Long and Summers (1986b).

[5] This specification was first suggested in this context by Robert Driskill and Steven Sheffrin (1986) to aid the analytical solution of the model. This simplification is unnecessary when numerical simulation methods are used, but was adopted by De Long and Summers. The implications of dropping this assumption are discussed in Section II.

[6] The assumption that g is the same in both equations (4) and (5) is somewhat arbitrary. The qualitative results are unchanged by allowing its value in (5) to lie between one-half and twice the value in (4). All that is necessary to obtain stabilizing price flexibility is that wages and prices respond at all to the current year's output.

TABLE 1—OUTPUT VARIANCE RESULTING FROM A UNIT DEMAND SHOCK

	I. ($A=1$, $\beta=0.0$) θ					II. ($A=1$, $\beta=2.5$) θ			
	1.0	0.9	0.8	0.7		1.0	0.9	0.8	0.7
g					g				
0.20	75	73	71	69	0.20	107	106	104	101
0.40	78	74	70	64	0.40	115	110	105	98
0.60	81	74	67	59	0.60	121	112	103	91
0.80	83	74	64	54	0.80	125	113	99	84
1.00	85	73	61	49	1.00	129	112	95	78
	III. ($A=2$, $\beta=0.0$) θ					IV. ($A=2$, $\beta=2.5$) θ			
	1.0	0.9	0.8	0.7		1.0	0.9	0.8	0.7
g					g				
0.20	46	45	43	40	0.20	83	80	77	73
0.40	49	45	40	35	0.40	90	83	74	65
0.60	51	44	38	31	0.60	95	82	69	56
0.80	52	43	35	27	0.80	99	81	64	49
1.00	54	42	32	23	1.00	101	78	59	42

Note: The entries in each cell denote 100 times the variance of output that results from a shock to output demand of 1.0 in the first period and which dies out gradually with a serial correlation coefficient of 0.5. In each panel, prices are least flexible in the top left-hand corner, the responsiveness of wages to output increases moving down the columns, and the proportion of contract-based wages decreases moving to the right along the rows.

neous price level response to an output shock and the smaller will be the (destabilizing) increase in anticipated inflation. De Long and Summers, however, model increased flexibility as an increase in g, the responsiveness of wages and prices to anticipated excess demand for goods. An increase in g therefore implies an increase in wage/price responsiveness. Whether increased wage and price flexibility of this nature is stabilizing will depend on all of the parameters of the model. Analysis of this model for a range of plausible parameters was carried out in De Long and Summers (1986b) and is also discussed below.

II. Results

In Tables 1–3, we show how the variance of output resulting from demand and supply shocks varies with changes in g, the output *responsiveness* of wages, and θ, the degree of *stickiness* as measured by the proportion of contract-based wages. Prices are least flexible by either measure in the top left-hand corner of the table, and most flexible in the bottom right-hand corner. The output responsiveness of wages increases moving down the column, and the proportion of contract-based wages decreases moving to the right along a row. The model is solved using the Ray Fair/Taylor algorithm (Fair and John B. Taylor, 1983)[7] with parameter values suggested by De Long and Summers (1986b).[8]

A. *Demand Shocks*

In Table 1, we show the variance of output that results from a serially correlated

---

[7] The PASCAL source code for the model solution is available from the author on request.

[8] De Long and Summers allow A to vary between 1.0 and 3.0, implying a real interest rate elasticity of spending between 0.08 and 0.25. The interest semi-elasticity of money demand is assumed to be around 2.5, implying an interest elasticity of money demand of 0.33 at an interest rate of 7.5 percent. To be consistent with De Long and Summers, we assume that b, the lag coefficient in the aggregate demand equation, is zero.

TABLE 2—OUTPUT VARIANCE RESULTING FROM A UNIT SUPPLY SHOCK

	I. ($A=1$; $\beta=0.0$) θ					II. ($A=1$, $\beta=2.5$) θ			
	1.0	0.9	0.8	0.7		1.0	0.9	0.8	0.7
g					g				
0.20	34	32	30	28	0.20	28	27	26	24
0.40	31	28	26	23	0.40	27	25	23	21
0.60	29	26	24	20	0.60	27	24	21	18
0.80	29	25	22	18	0.80	26	23	20	17
1.00	28	24	20	16	1.00	26	22	18	15
	III. ($A=2$, $\beta=0.0$) θ					IV. ($A=2$, $\beta=2.5$) θ			
	1.0	0.9	0.8	0.7		1.0	0.9	0.8	0.7
g					g				
0.20	76	72	67	61	0.20	80	75	70	64
0.40	71	64	56	49	0.40	78	69	60	52
0.60	68	59	49	41	0.60	77	64	53	43
0.80	67	55	44	34	0.80	76	60	47	37
1.00	66	52	40	30	1.00	76	57	43	32

Note: See Table 1.

shock to output demand.[9] The De Long and Summers' finding is illustrated in the first column of each panel of Table 1. As the output responsiveness of wages and prices increases, output variance increases.[10] By contrast, as was suggested in Section I, output variance unambiguously declines when the stickiness of wages decreases (θ falls).

Further, the De Long and Summers' result is weakened as the proportion of contract-based wages, θ, declines. For example, in the second column of the top left-hand panel, in which 90 percent of wages and prices are contract-based, increases in g decrease output variance over some range. What is more, the last two columns reveal that if no more than 80 percent of wages and prices are contract-based, increases in g *uniformly* decrease output variability—in direct contrast to the findings of De Long and Summers. Much the same findings emerge in the other three panels. Panels II and IV show that if either the money supply or demand elasticities are higher, output is more variable, because the attenuating movements of interest rates are smaller for higher elasticities. Panels III and IV reveal that a larger responsiveness of output to the real interest rate (A) leads to smaller output volatility, for the reverse reason.

B. *Supply Shocks*

It is also important to investigate how the results change if the shock is to aggregate supply, not demand. Table 2 displays the output variances that result from a one-time unit shock to the contract wage (u_t^s in equation (4)). In response to supply shocks, increased flexibility of *either* variety results in a lower-output variance for all parameter

---

[9] The magnitude of the initial shock is 1.0, and the serial correlation coefficient is 0.5, as assumed in De Long and Summers (1986b). The table displays the variances multiplied by 100.

[10] The finding that price flexibility can be destabilizing depends critically on the price level's being independent of the shock to the current year's output. The destabilizing price flexibility result is attenuated and may disappear if the price equation (6) allows that the price markup be related to output, perhaps because of cyclical sensitivity of raw materials or commodities prices. For example, if the coefficient on output in the price equation is 15 percent of g, the entries in the first column of the first panel of Table 1 become 71, 72, 71, 70, 68. With a coefficient greater than 15 percent of g (or 10 percent of g, if A were 2.0), increased price flexibility is unambiguously stabilizing.

TABLE 3—OUTPUT VARIANCE RESULTING FROM DEMAND AND SUPPLY SHOCKS
($A=1.0$, $\beta=0.0$)

A. Nominal GNP Fixed

	I. Demand Shock					II. Supply Shock			
	θ					θ			
g	1.0	0.9	0.8	0.7	g	1.0	0.9	0.8	0.7
0.20	0	0	0	0	0.20	104	95	85	75
0.40	0	0	0	0	0.40	81	71	62	52
0.60	0	0	0	0	0.60	71	60	50	41
0.80	0	0	0	0	0.80	65	53	43	34
1.00	0	0	0	0	1.00	60	48	38	28

B. Expected Nominal GNP Fixed

	III. Demand Shock					IV. Supply Shock			
	θ					θ			
g	1.0	0.9	0.8	0.7	g	1.0	0.9	0.8	0.7
0.20	51	49	46	43	0.20	79	71	62	54
0.40	51	47	42	37	0.40	56	49	42	35
0.60	51	45	39	32	0.60	46	39	32	26
0.80	51	43	35	38	0.80	40	33	27	22
1.00	51	41	33	25	1.00	36	29	23	18

Note: The demand shock is, as in Table 1, serially correlated, while the supply shock, as in Table 2, is nonrecurring.

values examined.[11] The explanation for this finding is simple. A positive wage/price shock decreases output contemporaneously, but the larger g is or the smaller θ is, the more this output decrease is attenuated by the decline in the real interest rate. Hence, more flexible prices cause output to be unambiguously more stable in response to supply shocks.

C. *Shocks with Nominal GNP Targeting*

It is possible that the choice of monetary rule is distorting the relationship of price flexibility and output stability. In Table 3 we examine the stability of output in response to demand and supply shocks in the presence of two types of nominal GNP targeting policies. In the top panels, monetary policy is assumed to successfully keep nominal GNP fixed. The left panel shows results for demand shocks and the right panel shows them for supply shocks. As is well known, aggregate demand policy can, in principle, exactly offset demand shocks so the variance of real output is zero for all parameter combinations. But we again see that in the case of supply shocks, increased flexibility of either variety (movements toward the top left of the table) decreases output variance.

The assumption that the monetary authority is capable of exactly controlling nominal GNP is extreme, so in the bottom two panels we examine the results assuming that the money supply at time $t-1$ is manipulated to maintain a constant expected value of GNP at time t. Of course, such a policy can no longer fully offset demand shocks, as the variances in the left panel (III) demonstrate. If wages and prices are completely contract-based ($\theta=1$), changes in g do not affect the output variance resulting from demand shocks because the change in current output is independent of g. But for any value of θ less than 1.0, increases in g uniformly decrease output variance. If the government attempts to stabilize GNP with an information lag, then decreased price flexibility will unambiguously reduce its success.[12]

---

[11] Taylor (1986b), Driskill and Sheffrin (1986), and De Long and Summers (1986b) have all shown that price flexibility is stabilizing in the De Long and Summers' framework.

III. Conclusion

This paper investigated the effects of increases in two types of price flexibility in the context of an extension of the Taylor (1980) model that contained De Long and Summers' model as a special case. The results can be simply summarized:

1) Increased flexibility, in the form of a decrease in the proportion of wages negotiated under contract, unambiguously decreases the variance of output resulting from both demand and supply disturbances.

2) Increased flexibility in the form of an increased responsiveness of wages and prices to expected future output may decrease the variance of output resulting from demand shocks, but only if the proportion of wages determined under contract is close to unity. Increased flexibility of this form will necessarily be stabilizing in response to supply shocks.

3) Increased flexibility of either variety is always stabilizing if monetary policy targets either actual nominal GNP or the expected value of GNP one period ahead.

Neither decreased price stickiness nor increased price responsiveness reduce the output variability that results from supply shocks, and can only decrease output variability resulting from demand shocks in a very limited set of circumstances. If the changes in the nature of price flexibility between the pre-1914 and postwar period are decreases in the proportion of spot-negotiated wages rather than a decreased output responsiveness of predetermined wages and prices, or if there is any contemporaneous effect of output on either prices or wages, then increased price stickiness cannot have contributed to the postwar increase in output stability. The explanation for increased stability must be sought elsewhere.

REFERENCES

Ambler, Steve and Phaneuf, Louis, "Price Flexibility Is Stabilizing After All," Université Du Québec à Montréal, Cahier no. 8721D, 1987.

De Long, J. Bradford and Summers, Lawrence H., (1986a) "The Changing Cyclical Variability of Economic Activity in the United States," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, 1986.

______, (1986b), "Is Increased Price Flexibility Stabilizing?," *American Economic Review*, December 1986, *76*, 1031–43.

Driskill, Robert A. and Sheffrin, Steven M., "Is Price Flexibility Destabilizing?," *American Economic Review*, September 1986, *76*, 802–07.

Fair, Ray and Taylor, John B., "Solution and Maximum Likelihood Estimation of Dynamic Non-Linear Rational Expectations Models," *Econometrica*, July 1983, *51*, 1169–85.

Romer, Christina, "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, February 1986, *94*, 1–37.

Taylor, John B., "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, *February* 1980, *88*, 1–23.

______, (1986a) "Improvements in Macroeconomic Stability: The Role of Wages and Prices," in *The American Business Cycle:* ...1986.

______, (1986b) "Reply to J. Bradford De Long and Lawrence H. Summers," in *The American Business Cycle:* ...1986.

---

[12] As long as the behavioral coefficients in the model are unchanged by the stabilization policy.

Is Increased Price Flexibility Stabilizing?: Reply

By J. BRADFORD DE LONG AND LAWRENCE H. SUMMERS*

Stephen King's comment on our papers (Bradford De Long and Lawrence Summers, 1986a, b) challenges our conclusion that over the range which is empirically relevant for economies like that of the United States, increases in price flexibility are likely to be destabilizing. He raises two main issues. First, King alleges that our results are highly sensitive to the particular modeling strategy we employed in considering changes in the degree of wage and price flexibility. Second, King argues that sufficiently prudent monetary policies could vitiate our results: increased wage and price flexibility would be stabilizing in the presence of "optimal" monetary policies. We treat these points in turn. Neither leads us to modify our views on the issues at hand.

I. Modeling Wage and Price Rigidity

In De Long and Summers (1986a) we explored the effects on economic stability of changes in the degree of wage flexibility within the context of John Taylor's (1979) model of overlapping contracts.[1] We reached similar conclusions—that increases in price flexibility could be destabilizing—regardless of whether we modeled decreased price stickiness by reduced contract lengths, increased sensitivity of wages to changes in output, or reduced amount of calendar time represented by a single period in the model. Nonetheless, King claims that our major result regarding the destabilizing effects of increased price flexibility is "not robust to... an alternative formulation of price stickiness."

King's preferred alternative formulation postulates that the economy is comprised of a contract sector, in which a fraction θ of all workers are located, and a "spot-market" sector, in which the remaining fraction of workers are located. In the spot-market sector nominal wages are proportional to the aggregate price level and depend positively on contemporaneous total output. King does not describe how the "spot" and "contract" sectors coexist. If an economy did contain a true spot market for labor, employment fluctuations could not be observed—workers laid off in the contract sector would simply present themselves in the spot-market sector's marketplace. Alternatively, it might be assumed that labor is immobile between the spot and contract sectors. In this case employment fluctuations are possible, but wages in the spot-market sector would not depend on *total* output.

There is a more profound difficulty with King's approach. In his spot sector wages are not predetermined over any time period, no matter how short. Wages move discontinuously and unexpectedly, and so they can adjust without ever causing any change in *ex ante* real interest rates. The channel through which price flexibility might be destabilizing—the impact of anticipated price changes on real interest rates—is simply assumed away. Bennett McCallum (1983) has already made the point that destabilizing inflation simply cannot occur if one ignores the process of price adjustment and simply postulates instantaneous market clearing. It is only a small leap to King's conclusion that increasing the share of the economy in which the problem of price adjustment is assumed away makes the economy more stable in the face of shocks.

To see that King's conclusions depend not on the reduction in the number of multi-period contracts but on the removal by as-

---

*Department of Economics, Harvard University, Cambridge, MA 02138. We would like to thank the Alfred P. Sloan and National Science Foundations under grant no. SES-8553244 for financial support, and Olivier Blanchard for helpful discussions.

[1] We would have reached very similar qualitative conclusions had we chosen Fischer or Calvo's overlapping contracts model. See De Long (1987b) and Binky Chadha (1987).

TABLE 1—BLOCK IV OF KING'S TABLE 2 WITH (4′) IN PLACE OF (4)

	$A=2.0,\ k=2.5$			
	θ			
	1.0	0.9	0.8	0.7
$g=0.2$:	103	106	109	113
0.4:	113	117	121	126
0.6:	119	124	129	134
0.8:	124	129	135	141
1.0:	128	134	140	146

TABLE 2—KING'S TABLE 2, BLOCK I, MINUS KING'S TABLE 4, BLOCK III

	$A=1.0,\ k=20.0$			
	θ			
	1.0	0.9	0.8	0.7
$g=0.2$:	23.4	24.2	24.7	25.2
0.4:	27.1	27.5	27.5	27.0
0.6:	29.8	29.6	28.5	26.7
0.8:	31.9	30.9	28.7	25.7
1.0:	33.7	31.5	28.2	24.3

sumption of the process of price adjustment as a potential disturbing factor, consider a variant of King's model in which instantaneous price-level jumps are not allowed. Replace King's equation (4) with the following (4′)

$$(4')\qquad W_t^s = E_{t-1}(p_t + g(y_t)).$$

This equation holds that the nominal wage is predetermined with respect to contemporaneous disturbances in the economy. It corresponds to the treatment of wage and price setting that is standard in Keynesian textbooks (for example, Frederic Mishkin, 1986, or Robert Hall and John Taylor, 1986) and which was employed by Thomas Sargent and Neil Wallace (1975) in their analysis of the effects of monetary policy.

Table 1 reports the results of redoing the calculations in King's Table 2 that were least favorable to our conclusions using equation (4′) rather than his (4). Increases both in the sensitivity of prices to output and in the spot-market sector's share of the economy are destabilizing.

There is a logically distinct point supporting our conclusion that price flexibility is likely to be destabilizing. We focused on the possibility that price flexibility could be destabilizing through its effect on real interest rates. An equally plausible line of argument suggests that price flexibility is destabilizing because of the financial strains it creates. Ben Bernanke (1983) makes a persuasive case that past unexpected price-level declines have an important effect on demand because of their impact on the solvency of the banking sector. If this debt-deflation channel is important in addition to or in place of the *ex ante* real interest rate, then King's major point is without force.

II. Monetary Policy

King makes the well-known point that properly set monetary policy can eliminate anticipated business fluctuations and leave only uncorrelated and unanticipated deviations of output from trend. In such an environment in which demand fluctuations never cause persistent output movements, King is correct that price flexibility would be stabilizing in our models if monetary policy were optimal. But in the real world demand fluctuations do cause persistent output fluctuations. King's appeal to an optimal monetary policy thus appears to be of little relevance in practice. Serially correlated fluctuations in output and employment are a ubiquitous feature of historical experience.

An interesting facet of King's analysis is that it implicitly suggests that flexible monetary policy and flexible wages might be complements. Flexible monetary policy is usually thought of as an alternative to flexible prices: if prices were more flexible, monetary policy could be more rigid. But in King's analysis increased price flexibility can make it more, not less important that monetary authorities take action to stabilize real demand.

This can be seen by examining our Table 2, which simply subtracts block III of King's Table 4 from block I of King's Table 2. This difference is a measure of the amount by which business cycle variance is reduced as a result of a move from an inappropriate to an optimal monetary policy rule. Over a range, as θ increases the effect on business cycle size of the shift to proper monetary policy increases. Increasing price flexibility raises the importance of aggregate demand management. If prices are more flexible, then the macroeconomic losses from relying not on nominal money stock but on price-level changes to stabilize the real money stock can be amplified: price-level changes affect the real interest rate, while nominal money stock changes do not.

III. Observations

It is clear that models can be constructed in which increases in price flexibility are destabilizing. At this point, the interesting issues are empirical. A variety of considerations suggest to us that destabilizing price flexibility is more than just a theoretical possibility.

First, the post-World War II period has seen fewer violent economic fluctuations and less "Walrasian" labor and product markets than did the pre-World War II period. As our (1986a) paper demonstrates, stabilization cannot be easily traced either to the effective use of countercyclical fiscal policy or to a lower variance of surprise changes in the money stock.[2]

Second, countries (like those in Latin America) where prices and wages are extremely flexible—where high inflation should have ironed all nominal rigidities out of the economy—appear experience business cycles just as volatile as other countries. Lawrence Summers and Sushil Wadhwani (in preparation) find no noticeable relation between the magnitude of a nation's business cycle and various measures of the extent of nominal wage rigidity even after controlling for the variance of nominal shocks and for a variety of other factors. It is noteworthy that in the post-World War II period Japan has had output fluctuations about trend much more volatile than the United States despite Japan's high degree of wage flexibility. De Long's (1987a) comparison of the French and British returns to gold in the 1920s suggests a similar conclusion. Although France had more flexible prices, it suffered a sharper recession as a consequence of the monetary austerity associated with returning to gold.

Third, the experience of the Great Depression suggests that halting deflation was a prerequisite for, not a consequence of the beginning of recovery. (Recovery began after, not before, the announcement of the NRA —a broad-based program to halt the decline in the general-price level.) In an accounting sense *all* of the Great Depression was due to a collapse in monetary velocity that cannot implausibly be traced to deflation: recall that real broad money balances (M2) did not decline between 1929 and 1932.

Fourth, our own attempts (1986b) to use var systems found that price shocks are positively associated with future output movements in systems with more than two variables.[3] The fact that decreases in prices Granger cause decreases in output fits naturally with the view that deflation is a cause as well as a consequence of deficient real aggregate demand.

This set of empirical considerations is hardly conclusive. On the other hand, despite its status as conventional wisdom, we are aware of no empirical demonstrations that greater price flexibility is associated with reduced output variability. At this point, it seems fair to shift the burden of proof to those who continue to regard this proposition as self-evident.

---

[2] Christina Romer's (1986) revisions of pre-World War II output data cannot be used to challenge the conclusion that, taken as a whole, the pre-World War II period saw larger business cycles than the post-World War II period has seen. Claims that there has been no stabilization of the U.S. economy assume that the Great Depression—the largest single piece of evidence that the pre-World War II economy was unstable—was a freak event that carries no information about the properties of the underlying structure generating business cycles before World War II.

[3] Similar results were found by Charles Calomiris and Glenn Hubbard (1985).

REFERENCES

Bernanke, Benjamin, "Nonmonetary Influences of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review*, March 1983, *73*, 257–76.

Calomiris, Charles and Hubbard, R. Glenn, "Price Adjustment, Credit Availability, and Economic Fluctuations: Evidence from the United States, 1879–1914," NBER Working Paper No. 1757, March 1985.

Chadha, Binky, "Is Increased Price Inflexibility Stabilizing? Some Analytical Results," New York: Columbia University, Xerox, 1987.

De Long, J. Bradford, "Returning to Gold in the 1920s," unpublished doctoral dissertation, Harvard University, Cambridge, 1987a.

______, "Drawbacks of Aggregate Price Flexibility," Cambridge: Harvard University, Xerox, 1987b.

______, **and Summers, Lawrence H.,** "Is Increased Price Flexibility Stabilizing?," *American Economic Review*, December 1986a, *76*, 1031–44.

______, **and Summers, Lawrence H.,** "The Changing Variability of Economic Activity in the United States," in Robert J. Gordon, ed., *The American Business Cycle*, Chicago: University of Chicago Press, 1986b.

Hall, Robert and Taylor, John, *Macroeconomics*, New York: W. W. Norton, 1986.

McCallum, Bennett, "The Liquidity Trap and the Pigou Effect," *Economica*, November 1983, *50*, 395–405.

Mishkin, Frederic, *The Economics of Money, Banking, and Financial Markets*, Boston: Little, Brown, 1986.

Romer, Christina, "New Estimates of Gross National Product in the United States," Cambridge: NBER Xerox, 1986.

Sargent, Thomas and Wallace, Neil, "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *Journal of Political Economy*, April 1975, *83*, 241–54.

Summers, Lawrence H. and Wadhwani, Sushil, "Wage Rigidity and Cyclical Variability," in preparation.

Taylor, John, "Staggered Wage Setting in a Macro Model," *American Economic Review Proceedings*, May 1979, *69*, 108–113.

Wadhwani, Sushil, "The Macroeconomic Implications of Profit Sharing," *Economic Journal*, 1986 Annual Supplement, *97S*, 171–83.

The Choice Among Medical Insurance Plans: Comment

By FRANK D. GIANFRANCESCO*

There are qualifications to the theoretical comparison between Health Maintenance Organizations (HMOs) and conventional insurance (CI), presented in a recent contribution to this *Review* by Yael Benjamini and Yaov Benjamini (1986). These qualifications make difficult the acceptance, as an unambiguous prediction of economic theory, the authors' conclusions regarding the relative attractiveness of the two methods of insurance to groups of individuals with heterogeneous demands.

The authors' arguments are summarized as follows. For homogeneous insureds, an HMO can resemble ideal medical insurance depending on the extent to which the prearranged medical care provided in each health state approximates the desired level for the homogeneous group. Under such conditions, an HMO would be superior to a CI plan in that it is free of the "moral hazard" that causes a welfare loss under the latter. However, if insureds are heterogeneous in terms of "income, tastes, or any other factor that affects...demand," a welfare loss will also result under the HMO. Because the quantity of medical care provided for a given health state is fixed for an HMO, it cannot satisfy the divergent medical care demands of all insureds in a heterogeneous group. In contrast, the greater flexibility afforded by CI enables insureds with divergent demands to more closely achieve their desired levels of medical care consumption. The welfare loss due to moral hazard is assumed to be little affected by divergent demands, and no other potential effects of demand heterogeneity on the desirability of CI are mentioned. The implication of the authors' analysis is that greater heterogeneity among insureds generally increases preferences for a CI plan over an HMO.[1]

The analysis presented by Benjamini and Benjamini (1986) contains two omissions that pertain to the effect of heterogeneous demand on the desirability of a CI plan. First, heterogeneous demand causes cross-subsidization under a CI plan over and above that due to the incidence of health state. Low demanders subsidize high demanders under CI, whereas, no such cross-subsidization exists under HMOs. This makes ambiguous the effect of heterogeneous demand on general preferences for the two methods of insurance. Second, there appears to be no theoretical basis for assuming that the size of the welfare loss under CI is virtually unaffected by heterogeneous demand. Heterogeneous demand can be shown to increase the welfare loss under a CI plan when this loss is more accurately measured. Thus, the effect of heterogeneous demand, at least in theory, is for this reason much more ambiguous than the Benjamini and Benjamini article implies.

The first point is demonstrated with the use of Figure 1. A single, precisely defined unhealthy state, *X*, and the divergent medical care demands of two representative individuals, A and B, are assumed. A and B are identical in terms of health risk but have differing demands for medical care because of differences in income, tastes, etc. Their respective demands for medical care under a specific CI plan are depicted in Figure 1. Zero administrative costs and a coinsurance rate of .2 are assumed. The price per unit of care is assigned a value of 10 and is equal to marginal cost, which is assumed constant. Under these constraints, individual A de-

---

*Senior Health Care Reimbursement Specialist, Abt Associates Inc., Washington, DC 20008.

[1] See Benjamini and Benjamini (1986): p. 222, 2nd para.; Section II, pp. 222–24; and p. 226, 3rd full para.

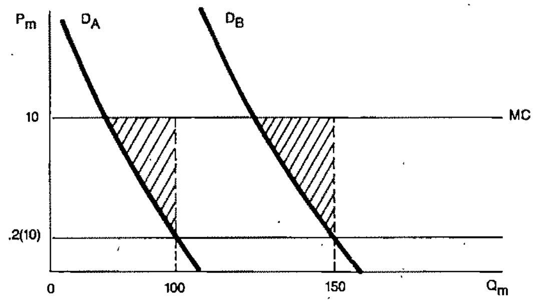


FIGURE 1

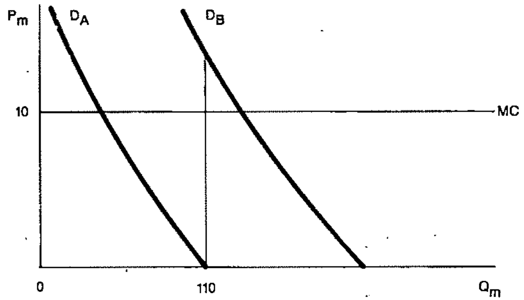


FIGURE 2

mands 100 units while individual B demands 150. Because of their differing demands, the insurance compensation received by each also differs, being 800 for individual A and 1200 for individual B. Assuming that the probability of unhealthy state X is .10, the premium is

$$.10((800+1200)/2)=100.$$

Thus, individual A pays a premium of 100 for an expected compensation of 80 while individual B pays the same premium for an expected compensation of 120. In effect, A has subsidized B's higher medical care consumption by 20.

The absence of the above effect in an HMO is demonstrated in Figure 2. All assumptions except that pertaining to the method of insurance are the same. Under the HMO both individuals are constrained to consume 110 units of care in the event of unhealthy state X. While this fixed amount exceeds individual A's desired level and falls short of B's, thereby creating a welfare loss, the inflexibility of the HMO also precludes any cross-subsidization. Both individuals receive in-kind insurance compensation valued at 110 and pay a premium of 110. For each insured under the HMO, the premium paid equals expected compensation irrespective of whether the insured group is comprised of individuals with homogeneous or heterogeneous demands.

While the cross-subsidization under a CI plan does not imply any welfare loss in the strict efficiency sense of the term, it does affect the desirability of the plan.[2] It will be less attractive for insureds with lower than the average demand and more attractive for those with greater than the average. This makes difficult the prediction that a stronger overall preference for CI will be associated with greater heterogeneity of demand.

General acceptance of the Benjamini-Benjamini (B-B, 1986) prediction is made more difficult when the effect of heterogeneous demand on the welfare loss from moral hazard is subjected to closer scrutiny. That heterogeneous demand can significantly increase this loss follows directly from the analysis of moral hazard first presented by Mark Pauly (1968). The welfare loss from moral hazard has two components. The first is due to the excess consumption of medical care caused by the price effect inherent in CI. The second component of welfare loss is due to the cost sharing required to moderate this effect. While cost sharing reduces excess consumption, it also reduces the risk-absorbing benefits of insurance. Thus, a tradeoff exists between the two components of welfare loss. The efficient level of cost sharing under CI is that which minimizes the total welfare loss.[3]

Though it is common for authors to separate the welfare effects of insurance coverage per se from those pertaining to the effects of insurance on resource allocation (i.e., moral hazard), this cannot be done here. The theoretical framework presented by B-B (1986) is used to predict consumer choice between HMOs and CI. The welfare effects of increasing or reducing coverage under CI certainly influence this choice and, therefore, cannot be ignored.

---

[2] Our inability to make interpersonal comparisons of utility precludes any demonstration of welfare loss as a consequence of the cross-subsidization.

[3] While B-B mention these two components of the welfare loss, they do not consider how the tradeoff between the two and therefore, how the size of the overall welfare loss, is affected by heterogeneous demand under conventional insurance.

It follows that the efficient level of cost sharing under CI cannot, except by extreme coincidence, be the same for all individuals. For each individual it would depend on his price elasticity of demand, his level of demand, and his degree of risk aversion. Thus, while the cost sharing offered by a CI plan could be efficient for a homogeneous group, it is unlikely to be for all insureds in a heterogeneous group. The following graphical exposition demonstrates this argument.

Let us again refer to Figure 1. For simplicity we have assumed that the slopes of D_A and D_B are identical. Thus, the price effect and the related welfare loss are the same for both individuals.[4] Though the welfare loss from excess consumption is the same for both individuals (because we have assumed it as such), the differing levels of demand suggest that the welfare loss from cost sharing is not. A coinsurance rate of 20 percent causes individual B to incur out-of-pocket expenses of 300 while A's are 200. If both individuals have the same degree of risk aversion, then B's welfare loss from a 20 percent cost sharing is greater than A's.[5] The tradeoff between the welfare loss from excess consumption and that from cost sharing is affected by the level of demand. B's higher level of demand means that each increment in cost sharing results in more uncovered medical expenses and related welfare loss than it does for A. This implies that B's efficient level of cost sharing is lower than A's. The total welfare loss would be less if A and B were separately insured under plans that offered the efficient cost-sharing levels for each.

Differences among insureds in terms of price effects and risk aversion provide additional ways in which heterogeneity can increase the welfare loss under a CI plan. For example, if B's demand curve were more steeply sloped than A's, implying smaller price effects, his tradeoff between the two components of welfare loss would differ even more from A's, and so would the efficient level of cost sharing. And if, in addition, B were more risk averse than A, the tradeoffs and efficient levels of cost sharing would be yet more divergent.

In summary, the B-B (1986) argument that greater heterogeneity among insureds is associated with a stronger general preference for conventional insurance over an HMO cannot be interpreted as an unambiguous prediction of economic theory. The analysis upon which their argument is based contains two omissions. First, heterogeneity causes cross-subsidization under a CI plan while this effect is absent from an HMO. This obfuscates the effect of heterogeneous demand on general preferences for either method of insurance. Second, the authors' assumption that heterogeneity does not affect the welfare loss from moral hazard has, upon closer examination, little basis in theory. Their analysis does not consider the fact that the size of the welfare loss under a CI plan is controlled by the level of cost sharing and that the efficient level would very likely differ among heterogeneous insureds.

The empirical results presented by B-B (1986) do show more homogeneity among insureds with an HMO in comparison to a CI plan that was studied. These results, however, were for a single observation; one HMO and one CI plan were compared for one population (6,168 University of Pennsylvania employees). While it may be the case that, for the sample studied, the negative effects of heterogeneity were stronger for the HMO, it cannot be concluded that this is generally true. Further, even if the B-B (1986) findings are generalizable, there is another explanation, as the authors themselves indicate, for observing more homogeneous individuals within HMOs. A review of the literature on HMO enrollment by Sylvester Berki and Marie Ashcraft (1980) and later work by Joan Buchanan and Shan Cretin (1986) conclude that younger individuals, who do not have established patient-physician relationships, are more likely to enroll in HMOs. While Benjamini and Benjamini argue, citing an earlier logit analysis, that age is insufficient to totally explain the choice between HMOs and CI (and we certainly do not disagree), this point seems irrelevant to their findings. Younger age may be sufficient to explain the tighter variances for age, income, and education (i.e., the greater homogeneity) observed among the HMO enrollees in their sample.

---

[4] B-B (1986) make the same implicit assumption in their graphical analysis. Also like the authors, we have assigned a welfare loss to all additional medical care consumption under CI. A more precise rendition of welfare loss would compare medical care consumption under CI to that which would exist under ideal insurance. See, for example, Kenneth Arrow (1973) and Frank Gianfrancesco (1978).

[5] We, of course, make no interpersonal comparisons of utility between the two individuals and assume that any reduction in insurance coverage is equally felt.

REFERENCES

Arrow, Kenneth J., *Optimal Insurance and Generalized Deductibles*, A Rand report prepared for the Office of Economic Opportunity, R-1108-OEO, February 1973.

Berki, Sylvester E. and Ashcroft, Marie L. F., "HMO Enrollment: Who Joins What and Why: A Review of the Literature," *Milbank Memorial Quarterly/Health and Society*, 1980, *58*, 588–625.

Buchanan, Joan and Cretin, Shan, *Fee-for-Service Health Care Expenditures: Evidence of Selection Effects Among Subscribers Who Choose HMOs*, A Rand study funded by the U.S. Department of Health and Human Services, R-3341-HHS, March 1986.

Benjamini, Yael and Benjamini, Yoav, "The Choice Among Medical Insurance Plans," *American Economic Review*, March 1986, *76*, 221–27.

Gianfrancesco, Frank D., "Insurance and Medical Care Expenditures: An Analysis of the Optimal Relationship," *Eastern Economic Journal*, July/October 1978, 225–34.

Pauly, Mark V. "The Economics of Moral Hazard: Comment," *American Economic Review*, June 1968, 531–37.

The Choice Among Medical Insurance Plans: Reply

By YAEL BENJAMINI AND YOAV BENJAMINI*

In our 1986 paper, "The Choice Among Medical Insurance Plans," we first showed that for a homogeneous group of individuals an HMO plan is welfare superior to a CI plan. We then relaxed the assumption of homogeneity in our analysis, and demonstrated that "when heterogeneity is sufficiently strong a CI plan may be better." Frank Gianfrancesco (1988) raises two qualifications that are supposed to "make difficult the acceptance, as an unambiguous prediction of economic theory," (of this second result). We clearly realized, and stated, that we were not making unambiguous theoretical predictions, so in some sense Gianfrancesco is breaking into a house which has an open door. However, we did support this second result by analyzing data on groups of employees enrolled in CI and HMO, and found the latter group to be much more homogeneous in income, age, and education. This was true both in absolute terms and relative to the levels of the variables. Because of these results, we find it difficult to believe that the large differences in homogeneity between the two groups are all due to a difference of less than a half standard deviation in age—as suggested by Gianfrancesco. Furthermore the phenomenon of younger age in HMO insureds is coupled, in the sources quoted by Gianfrancesco, with lower income. We certainly do not see this in our results.

As to the qualifications themselves: regarding the cross-subsidization issue, Gianfrancesco takes a simple example and shows that under CI, an insured demanding less in a specific state of health, subsidizes the one demanding more when in that state of health. Let us ignore the fact that this does not imply efficiency loss in the strict sense of the term. An individual may be a high demander in one state of health while being a low demander in a second state of health. Therefore, it is still not clear, though is assumed so by Gianfrancesco, that a single individual will emerge as high demander when all possible states of health are accounted for. Thus it is not clear at all that, *ex ante*, cross-subsidization exists at any significant level.

The qualification regarding the size of welfare loss due to moral hazard is, in our view, a more important one. The separability of the utility function appears in many works, including those comparing insurance plans. Gianfrancesco shows that when both homogeneity and separability are relaxed the effect of heterogeneity may work in the direction that reduces the desirability of the CI plan. The next theoretical contribution naturally called for is an analysis which is set up to understand the conditions under which one effect is stronger than the other.

Ultimately this controversy should be resolved by empirical evidence. This may be accomplished by including for each person, a measure of his distance from the center of his group, as one of the variables to be used for explaining his choice. In the population we studied the group formed in an HMO was more homogeneous, but the question of generalizability can be answered only by new data. We hope this discussion draws attention to the importance of the heterogeneity issue in the health insurance marketplace, and will lead to additional studies.

REFERENCES

Benjamini, Yael and Benjamini, Yoav, "The Choice Among Medical Insurance Plans," *American Economic Review*, March 1986, *76*, 221–27.

Gianfrancesco, Frank D., "The Choice Among Medical Insurance Plans: Comment," *American Economic Review*, March 1988, *78*, 277–80.

---

*Department of Economics and Department of Statistics, respectively, Tel Aviv University, Ramat Aviv 69978, Tel Aviv, Israel.

The Disinterest in Deregulation: Comment

By JOE A. BELL*

In an article in this *Review*, R. E. McCormick, W. F. Shughart II, and R. D. Tollison (1984), hereafter MST, argue that there is a lack of interest in deregulation because the costs of rent seeking to obtain a monopoly position are sunk, and cannot be recovered. In their words there are, "...few people who stand to gain from regulatory reform," and consequently not much political support for deregulation. This conclusion is based on a misunderstanding of sunk costs and an assumption of asymmetry in resource mobility. MST assume that following regulation resources become specialized in new uses, but cannot reverse that specialization if the regulation is removed. It will be shown that without this asymmetry the analysis of MST cannot be supported.

The basic argument of MST is illustrated by their transformation curve diagram (MST's Figure 2). Regulation of industry A, with no rent seeking, will move the economy from the Pareto point 1 to a non-Pareto point 2. MST label the inefficiencies resulting from the movement to point 2 as Harberger costs (i.e., deadweight losses). Resources used in rent seeking will move the transformation curve for A and B from P_1 to P_2, and production to a point such as 2′. The inefficiencies of this lower transformation curve are labeled Tullock (rent seeking) costs.

According to MST, deregulation of industry A only permits the recovery of the Harberger costs and will move the economy to a point such as 3 on the new transformation curve (P_2). The economy cannot recover the Tullock costs (i.e., shift to the original transformation curve) because some resources have become specialized in rent seeking and can no longer be used to produce goods A and B. These Tullock costs are sunk and the economy will operate on the lower transformation curve indefinitely.

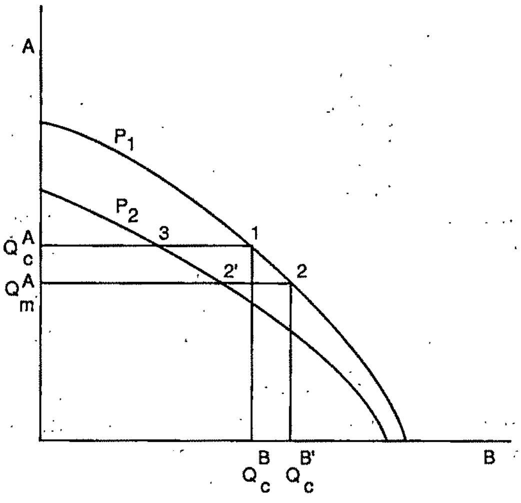


FIGURE 1. MST'S FIGURE 2

The fact that Tullock costs are sunk has nothing to do with resources becoming specialized in rent seeking. Any costs that represent past resource use are sunk, including Tullock costs. But by the same logic, Harberger costs are also sunk. Deregulation cannot recover either the Harberger or Tullock costs, but can only eliminate the recurrence of those costs in future periods.

The conclusions of MST rest on the asymmetric treatment of resource mobility, depending on whether regulation shifts resources away from the regulated industry A to industry B, or from A to rent seeking. In their model resources that move from A to B can, upon deregulation, return again to the production of A and end further Harberger costs. They assume, however, that resources which have become specialized in rent seeking cannot, following deregulation, return to the production of A, thus preventing the economy from regaining the original transformation curve (P_1). It is as though the capital value of resources (not just the services they provide in a given year) is destroyed by their application to rent-seeking activities.

---

*Department of Economics, Southwest Missouri State University, Springfield, MO 65804.

In a one-period analysis resource immobility may be a useful assumption. Even so, the immobility would also apply to resources lost to Harberger costs. In a multiperiod analysis, there seems to be no reason to view resources as permanently immobile and then only those resources used for rent seeking. If all resources are mobile, deregulation will move the economy back to the original point(1), eliminating both Harberger and Tullock costs in future periods.

To summarize, MST's analysis rests entirely upon the unexplained asymmetries in resource mobility. There may be reasons for the asymmetric treatment, but MST failed to provide them.

REFERENCES

McCormick, Robert E., Shughart, William F. II, and Tollison, Robert D., "The Disinterest in Deregulation," *American Economic Review*, December 1984, *74*, 1075–79.

The Disinterest in Deregulation: Reply

By ROBERT E. MCCORMICK, WILLIAM F. SHUGHART II, AND ROBERT D. TOLLISON*

Our paper in this *Review* (1984) aroused a controversy we did not anticipate, but no one has yet convinced us that our basic point is wrong. Joe Bell (1988, p. 282) now makes the assertion that our conclusion rests entirely upon "unexplained asymmetries in resource mobility."

This seems to us a very curious assertion indeed. The main problem with Bell's analysis is that he treats investment in capital assets as perfectly malleable. A lawyer trained to argue rate cases is not perfectly suited for other jobs when electric utility deregulation occurs. *The time and effort spent by the lawyer to acquire the requisite skills are irretrievably lost.* This is the point of our paper. Even though future labor can be supplied by this lawyer after deregulation, it definitionally has a lower value. Lawyering before a regulatory commission is a specialized input. When the demand for these services falls, the capital value of the intensive and extensive investments vanish. No asymmetry is implied or required.

More generally, capital consumed in using the political process to secure a wealth transfer—the resources devoted to organizing coalitions, investing in lobbying activities, contributing to political campaigns, advertising a point of view, acquiring the stock of human capital necessary for dealing with regulatory bureaucracies, and so on— reduces the wealth of society in opportunity-cost terms. For the reasons just stated, the cost of obtaining additional output in the regulated industry following deregulation is higher.

We also pointed out that the increase in marginal costs due to what Bell calls "resource immobility" may only be a short-term phenomenon: "Over time, resources in the [deregulated] industry will adapt to the new competitive environment, and new resources coming into the industry will embody the requisite skills for working in a competitive rather than a government-sponsored sector." Thus, after the relevant adjustment period, marginal costs in the deregulated sector may decline, but it is simply wrong to suppose that deregulation enables the capital *value* of resources specializing in rent-seeking activities to be used again in the production of goods. Anyway, in present-value terms it does not take very long for this adjustment process to impose significant costs on the economy. This point is covered in fn. 7 in our original paper.

REFERENCES

Bell, Joe A., "The Disinterest in Deregulation: Comment," *American Economic Review*, March 1988, *78*, 282–83.

McCormick, Robert E., Shughart, William F. II, and Tollison, Robert D., "The Disinterest in Deregulation," *American Economic Review*, December 1984, *74*, 1075–79.

---

*McCormick: Department of Economics, Clemson University, Clemson, SC 29634; Shughart and Tollison: Center for Study of Public Choice, George Mason University, Fairfax, VA 22030.

Price Leadership and Welfare Losses in U.S. Manufacturing: Comment

By VAUGHAN DICKSON*

In the September 1986 issue of this *Review*, Micha Gisser estimates welfare losses in U.S. manufacturing. He uses a price leadership model and concludes (p. 762) that maximum welfare losses are 1.823 percent of GNP when price leaders collude and 0.114 percent when they act independently as Cournot firms. This note maintains these numbers underestimate welfare losses substantially in the case of collusion among leaders, and by a smaller amount otherwise. (Gisser's conclusion in favor of independent behavior is not challenged.) This bias occurs because Gisser calibrates his model so that the concentration numbers taken from the *Census of Manufactures* describe the industry at the imaginary competitive equilibrium rather than at the real noncompetitive equilibrium. The result is that Gisser's welfare losses describe an economy with lower concentration ratios than actually exist in the United States. A recalibrated model correcting for this problem is introduced and new estimates of welfare loss, confirming the downward bias, are obtained. The estimates are obtained using more accurate integral calculus rather than numeric methods. Last, since price leadership models have unequal marginal cost between leading and fringe firms, an estimate of the portion of welfare loss due to unequal marginal cost is provided.

I. The Model Recalibrated

Gisser's model is summarized by the following equations:

$$D_T = 100/P^n, \tag{1}$$

$$S_0 = (100 - C)P^\varepsilon, \tag{2}$$

$$S_L = (CMC^\varepsilon), \tag{3}$$

$$E = 100/P^n - (100 - C)P^\varepsilon. \tag{4}$$

D_T, S_0, S_L, and E stand respectively for total industry demand, the supply of price takers, the production of the leaders, and the excess demand curve confronting leaders. P, ε, C, and n represent price, price elasticity of supply, the four-firm concentration ratio, and the negative of the industry price elasticity of demand. MC is the marginal cost of leading firms.

To start the model, Gisser assumes the competitive price is \$1. This implies that competitive output is 100 units and that the percentage of output supplied by the four largest firms in competitive equilibrium is C. To compute welfare loss, Gisser considers (4) and a marginal revenue equals marginal cost equation for the leaders and solves for noncompetitive price (P_m) and output of the leaders, E_m. The welfare loss triangle for the leaders is then simply half their markup times ($E_m - C$). (See Figure 1(b).) The "real" welfare loss is this number blown up by the ratio of *Census* industry revenue to simulated revenue.

It should be clear from the above that the noncompetitive equilibrium concentration ratio is $E_m/[100/P_m^n]$, which is less than the value of C supplied from the *Census*. This difference, which depends on n, ε, and the markup, is important because after assumptions about n, ε, and oligopoly behavior are made, the sole empirical basis for welfare losses is the concentration ratios and industry revenue.

To recalibrate the model, equations (1) to (4) are adjusted as follows:

$$D_T = A/p^n, \tag{1a}$$

$$S_0 = BP^\varepsilon, \tag{2a}$$

$$S_L = KMC^\varepsilon, \tag{3a}$$

$$E = A/P^n - BP^\varepsilon. \tag{4a}$$

A, B, and K are demand and supply shifters. To restart the model, assume industry output in the *oligopoly equilibrium* is 100 units,

---

*Department of Economics, University of New Brunswick, P.O. Box 4400, Fredericton, N.B., Canada, E3B 5A3.

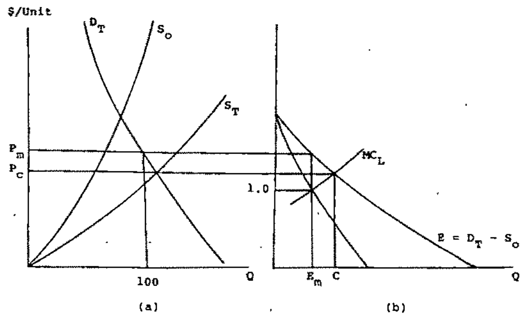


FIGURE 1. A PRICE LEADERSHIP MODEL

TABLE 1—DEADWEIGHT LOSS AS A PERCENTAGE OF GNP: 1977

	$n=1$	
ε	Collusion	Cournot
1	4.044	.126
	(1.823)	(.114)
10	1.601	.090
	(.294)	(.049)

Note: Numbers in parentheses are from Gisser (Table 2, p. 761).

the marginal cost of leading firms at the noncompetitive equilibrium is $1, and define CR as the *proportion* of industry output of the four leading firms at the *noncompetitive equilibrium*. From this it follows that at the noncompetitive equilibrium

$$(1-CR)100 = BP^{\varepsilon}, \tag{2b}$$

$$S_L = (100\,CR)MC^{\varepsilon}, \tag{3b}$$

$$100\,CR = A/P^{n} - BP^{\varepsilon}. \tag{4b}$$

Last, since n_E, the negative of the price elasticity of the net demand curve is $(n+(1-CR)\varepsilon)/CR$ and since the Lerner index for a leading firm, $((P_m-1)/P_m)$, will equal $1/n_E$ when leading firms collude and one-quarter of this when leading firms are Cournovian, it follows that[1]

$$P_m = 1/\left(1-[(CR/M)/(n+(1-CR)\varepsilon)]\right), \tag{5}$$

where $M=4$ with Cournot behavior and 1 with full collusion among leaders.

Given these preliminaries, inspection of Figure 1(b) reveals that the deadweight loss can be calculated as

$$L = \int_{P_c}^{P_m}(A/P^{n} - BP^{\varepsilon})\,dP + \int_{1}^{P_c}(100\,CR)MC^{\varepsilon}\,dMC - (P_m-1)100\,CR. \tag{6}$$

L may be computed for any industry if P_m, P_c, A, and B are known.[2] For given n, M, and ε, and for CR taken from the *Census*, P_m may be found using equation (5). B is found using equation (2b) and the already determined P_m. A is found using equation (4b) after substituting the solved values of B and P_m. Last, the competitive price, P_c, is found by setting leaders' marginal cost equal to E, (i.e., $100\,CRP^{\varepsilon} = A/P^{n} - BP^{\varepsilon}$) and solving for P.

II. Results and Conclusions

Equation (6) was used to estimate deadweight loss for Gisser's 445 industry sample for n and the range of ε considered most acceptable by Gisser, that is, $n=1$, $\varepsilon=1$, and 10. The results, together with Gisser's, are shown in Table 1 as a percentage of GNP. Under the assumption of collusion these results are considerably larger than Gisser's, the relative discrepancy being greatest when $\varepsilon=10$. The results are closer to Gisser's under the assumption of independent behavior. This is because CR values differ less between competitive and Cournot than between competitive and collusive positions. For example, the sample average for *CR* is 38 percent. If collusion is assumed the average *CR* implied by competitive equilibrium is 46, if $\varepsilon = 1$, and 50, if $\varepsilon = 10$. While with Cournot behavior the mean *CR* implied by competitive equilibrium is 40 and 42 for $\varepsilon = 1$ and 10, respectively.[3]

---

[1] The expression for n_E is found by differentiating $E = D_T - S_0$ with respect to price and converting to elasticities. The factor one-quarter assumes each leading firm is the same size. With Cournot behavior, each firm perceives the elasticity to be $4n_E$ because each assumes rivals will maintain output in the face of its output changes. The Lerner index for colluding dominant firms is derived in T. R. Saving (1970), and for Cournot leading firms in J. C. Hause (1977, p. 86).

[2] To convert to real losses, L is inflated by industry shipments divided by simulated revenue, $100P_m$.

Finally it should be noted that part of the welfare loss in price leadership models occurs because industry output is not supplied by firms with equal marginal cost. Fringe firms have higher marginal costs than leading firms. These estimates are therefore in a sense not comparable with estimates assuming equal marginal cost across firms. One can determine, however, what the welfare loss would be for a *given* P_m if marginal cost was equalized across all firms. This can be done by finding the deadweight loss triangle bordered by the total industry demand and the industry marginal cost curve, $S_T = S_0 + S_L$. (See Figure 1(a).) This is what Dean Worcester (1975) proposed for the deadweight loss measure in dominant firms models. Everything needed to evaluate this triangle is known, except for industry marginal cost when industry output is 100. This is easily found by setting $S_T = (100CR + B)P^{\varepsilon}$ to 100 and solving for P. Using methods similar to that outlined in (6), the welfare loss for the sample is 3.201 ($\varepsilon = 1$) and 1.087 ($\varepsilon = 10$) percent under collusion, and is .085 ($\varepsilon = 1$) and .044 ($\varepsilon = 10$) percent under Cournot behavior.[4] In comparison with Table 1, these figures tell us that while a majority of the losses can be apportioned to discrepancies between price and marginal cost, a significant amount is an efficiencies loss associated with the increased production of high-cost fringe firms.

REFERENCES

Gisser, Micha, "Price Leadership and Welfare Losses in U.S. Manufacturing," *American Economic Review*, September 1986, *76*, 756–67.

Hause, J. C., "The Measurement of Concentrated Industrial Structure and the Size Distribution of Firms," *Annals of Economic and Social Measurement*, Winter 1977, *6*, 73–107.

Saving, T. R., "Concentration Ratios and the Degree of Monopoly," *International Economic Review*, February 1970, *11*, 139–46.

Worcester, Dean, "On Monopoly Welfare Losses: Comment," *American Economic Review*, December 1975, *65*, 1015–23.

U.S. Bureau of the Census, U.S. Department of Commerce, *Census of Manufactures*, Washington: USGPO, 1977.

---

[3] The influence of *CR* on welfare loss is more than proportional since, as shown by Gisser (fn. 3, p. 758), welfare loss is quadratically linked to the Lerner index and hence *CR*. One might also ask what happens to Gisser's test of independent behavior with this adjusted model? Gisser's test was replicated with the numbers of this model with a slightly stronger result, that is, collusion could not be inferred until ε was about 40. At this ε, the welfare loss is 0.353 percent of GNP.

[4] An appendix, available from the author, illustrates in detail how this welfare loss and the first welfare loss based on equation (6) is computed for a representative industry, Motor Vehicle Parts, SIC 3714.

Price Leadership and Welfare Losses in U.S. Manufacturing: Reply

By MICHA GISSER*

Vaughan Dickson claims that I biased my results because (a) I used numeric methods rather than integral calculus, and (b) I calibrated my model so that the concentration numbers taken from the *Census of Manufactures* describe the industry at the imaginary competitive equilibrium rather than at the real noncompetitive equilibrium.

I. Numeric Methods vs. Integral Calculus

I calculated the noncompetitive price numerically because given the nature of the excess demand curve I could not calculate it analytically. However, in response to Dickson's suggestion, I have recalculated the losses by replacing the crude method of half-base times height (of the triangle) by analytic integration. The results for Dickson's selection of elasticities are presented in Table 1.

It is interesting to note that for $|\eta| = \varepsilon = 1$ the crude method resulted in upper-biased estimates of the loss. For $|\eta| = 1$ and $\varepsilon = 10$, the crude method is slightly downward biased.

TABLE 1—DEADWEIGHT LOSS AS A PERCENTAGE OF GNP: 1977

ε	Method of Calculation	$\eta = -1$	
		Collusion	Cournot
1	Crude[a]	1.823	0.114
	Integral	1.722	0.113
10	Crude[a]	0.294	0.049
	Integral	0.302	0.050

[a] Crude means half of the base of the triangle times its height.

II. Recalibrating the Model

I agree with Dickson that concentration ratios taken from the *Census of Manufactures* should describe the industry at the real equilibrium rather than the imaginary competitive equilibrium. In my study (1986) I argued that the real noncompetitive equilibrium is best approximated by a price leadership model in which the leaders follow a Cournot mode. My argument was confirmed empirically, and furthermore Dickson stresses that he does not challenge my conclusion in favor of independent behavior of the leaders. It follows that the proper adjustment is to assume that the concentration ratios taken from the *Census of Manufactures* reflect a Cournot mode followed by the leaders. As illustrated in an unpublished Appendix,[1] this once-and-for-all adjustment results in slightly higher concentration ratios associated with the imaginary competitive world. These new concentration ratios are as close as they can be to the real world. Dickson contradicts himself when on one hand he does not challenge my conclusion in favor of independent behavior, and on the other hand he rushes to calculate a second set of concentration ratios under the assumption that in real life the leaders collude. In my opinion, the only purpose of the collusion exercise is to set an upper limit to the welfare loss that would be created if the leaders were to collude.

The results of recalculating the losses after adjusting the concentration ratios are con-

---

*Department of Economics, University of New Mexico, 1915 Roma, N.E. Albuquerque, NM 87131.

[1] I will send a copy of the Appendix upon receiving a written request.

TABLE 2—DEADWEIGHT LOSS AS A PERCENTAGE OF GNP: 1977, CALCULATED FOR TWO SETS OF CONCENTRATION RATIOS

		Collusion Model Calibration:		Cournot Model Calibration:	
η	ε	Competitive Equilibrium[a]	Cournot Equilibrium[b]	Competitive Equilibrium[a]	Cournot Equilibrium[b]
−1	1	1.823	1.978	0.114	0.125
−1	10	0.294	0.473	0.049	0.087
−10	1	0.040	0.043	0.0025[c]	0.0027[c]

[a] Concentration ratios are assumed to reflect a competitive equilibrium (Gisser, 1986).

[b] Concentration ratios are assumed to reflect an equilibrium in which the 4 leaders adopt a Cournot mode of behavior. Also, the deadweight loss is estimated by applying integration methods, rather than measuring areas of triangles.

[c] In the text (Table 2, 1986) I reported a rounded result of 0.3.

tained in Table 2. The upper limit of the deadweight loss is now estimated at 0.125 instead of 0.114 percent of GNP. Under the unlikely assumption of collusion, the upper limit of the deadweight loss is revised from 1.823 to 1.978.

Finally, the last row in Table 2 demonstrates that the combination $\eta = -10, \varepsilon = 1$, should not be omitted from Dickson's calculations.

REFERENCES

Gisser, Micha, "Price Leadership and Welfare Losses in U.S. Manufacturing," *American Economic Review*, September 1986, *76*, 756–67.

Dickson, Vaughan, "Price Leadership and Welfare Losses in U.S. Manufacturing: Reply," *American Economic Review*, March 1988, *78*, 285–87.

Advertising and Economic Welfare: Comment

By ODED HOCHMAN AND ISRAEL LUSKI*

In a recent note, Len Nichols (1985), using the advertising model developed by George Stigler and Gary Becker (S-B) in their seminal work (1977), evaluates the efficiency of a market economy with advertising. However, his analysis is not entirely accurate. In this paper we remedy these inaccuracies and present a complete and correct analysis of the efficiency conditions in S-B's economy with advertising and identify the market allocations which fulfill these conditions.

In addition, we analyze S-B's model to derive results, beyond what Nichols and S-B did, and thus bring forward more fully the economy's market behavior.

The main results of our comment: A perfectly competitive commodity market with advertising constitutes a first-best solution and the monopolistic competition allocation is inefficient.

We show that a positive relationship exists between the firm's level of advertising, the firm's product output, and the firm's commodity output. This relationship is determined entirely by production characteristics, and has nothing to do with demand. Changes in demand affect advertising only due to changing the equilibrium output and with it the quantity of advertising.

We show, also, that in a perfectly competitive commodity market where the price of the commodity is constant the price of the product observable in the market may vary from firm to firm. The larger the firm, the higher its product price, its level of advertising, and its product output. These distinctive market features, which until now were used to characterize monopolistic competition, are shown here to typify perfect competition in the commodity market, which is also Pareto efficient.

Furthermore we show that the more concentrated a competitive industry is and the larger the firms in it—the more advertising is used and output produced per firm, than would have been in the case of firms in a less concentrated industry. Less product and more advertising per unit of commodity are used and the product prices are higher in a more concentrated industry even though it is perfectly competitive and operates efficiently.

Thus Stigler and Becker have unknowingly provided economic theory with a model which restores the efficiency properties to the invisible hand of retail markets, which this invisible hand was believed not to possess, since the 1930s—a decade during which Edward Chamberlin and Joan Robinson developed the theory of monopolistic competition to explain the operation of exactly such retail markets with advertising.

Analysis of the efficiency conditions and market allocations follows this section. Then we deal with the characteristics of different market allocations.

I. Efficiency and Market Allocations

We follow S-B's model and notations. We assume a population with homogeneous tastes, that is, all individual households have the same utility function $u(z, y)$, where z is a commodity consumed by the household and produced by it via a household production function using as inputs the product x, purchased in the market, and advertising A provided to all consumers by the producers of the particular brand of x being consumed. Let x and z designate the quantities of the product and the commodity, respectively, consumed by a household and X and Z the quantities produced by an individual firm. The quantity of advertising A produced by the firm is also the quantity consumed by each of its customers. Thus, as in S-B, equation (15), the household production function

---

*Department of Economics and The Monaster Economic Research Center, Ben-Gurion University, Beer-Sheva, Israel.

is

$$z = g(A)x. \tag{1}$$

There are several brands of z in the market, the product x of each brand being produced by its own firm which also purchases the required advertising and provides it free of charge to its own consumers. However, the consumers do not prefer one brand of z over another, all other conditions being identical. The variable y stands for a second consumption good provided by a competitive industry and consumed without advertising.

Consider two customers, i and k, of a given brand of z. Then, as in the neoclassical welfare literature,

$$\text{RCS}^i_{z,y} = \text{RCS}^k_{z,y}, \qquad \text{for all } i \neq k \tag{2}$$

is a Pareto efficiency condition. Here $\text{RCS}^i_{z,y}$ stands for the rate of substitution in consumption between z and y by household i.

If (2) is not fulfilled, the two individuals could exchange z with y between themselves, and by doing so raise the utility levels of both. Therefore (2) is necessary for Pareto efficiency. Since the two individuals consume the same brand, both are exposed to the same advertising and a transfer of z from one to the other involves only a transfer of x. Hence (2) also implies

$$\text{RCS}^i_{x,y} = \text{RCS}^k_{x,y}. \tag{3}$$

Indeed (3) follows from (2) upon noting that advertising A is the same for the two individuals, and that

$$\partial u/\partial x = g(A)(du/dz). \tag{4}$$

It should be noted again that relations (2) and (3) are efficiency conditions only if the two individuals are consuming the same brand of z. If they are consuming different brands, no such substitution in the margin of z by y can take place.

Let P_x be the market price of x and π_z the derived price of z. Then

$$P_x \cdot x = \pi_z \cdot z = \pi_z g(A) \cdot x, \tag{5}$$

which in turn implies

$$\pi_z = P_x / g(A). \tag{6}$$

In competition, utility-maximizing individuals equate marginal rate of substitution in consumption to the price ratio. Thus,

$$(\partial u/\partial x)/(\partial u/\partial y) = P_x/P_y, \tag{7}$$

and since P_x/P_y is identical to all individuals purchasing the same brand, (7) implies (3), and with it (2).

Consider now the production of z. The quantity of Z produced by the firm can be changed in the margin either by changing X or by changing A. Let L be a production factor, say labor, used in the production of both A and X. Then, to increase Z we can allocate an additional unit of L either to increase the production of X or to raise the level of A. If we allocated an additional unit of L to the production of A, and keep X unchanged, the increase in Z, which can be termed the marginal product of labor in the production of z through A, is[1]

$$(dz/dL)_{dx=0} = MP_L(A) \cdot g'(A) \cdot X, \tag{8a}$$

where $MP_L(A)$ stands for the marginal product of L in the production of A. If we allocated an additional unit of L to the production of X, and kept A unchanged, the increase in Z, termed the marginal product of L in the production of Z through X, will be

$$(dz/dL)_{dA=0} = g(A) MP_L(X). \tag{8b}$$

A Pareto efficiency condition is, that the firms in the economy operate so that the marginal products of L in the production of z through x and through A, are the same. In mathematical notation the above condition is

$$(dz/dL)_{dA=0} = (dz/dL)_{dx=0}, \tag{9}$$

---

[1] Equations (8a) and (8b) are obtained from equation (1) by differentiation.

and after substituting (8a) and (8b) into (9) we obtain

$$(10)\quad g(A)MP_L(X) = MP_L(A)g'(A)\cdot X.$$

That (9) and therefore (10) are efficiency conditions we prove by showing that if equality does not hold in (9) and (10) an inefficiency exists. Let us assume, as a working hypothesis, that inequality holds in (9) and (10), say the left-hand side (LHS) of (9) and (10) is larger than their respective right-hand sides (RHS). Then by shifting a unit of L from the production of A to the production of x, we change the total amount of produced z by the (positive by assumption) difference between the LHS and the RHS of (10) and thus increase the total amount of z without employing any additional inputs. This in turn implies that when inequality holds in (10) (and in (9)), we are not at Pareto optimum, which proves that equality in (9) and (10) is a necessary condition for such an optimum.

The left- and right-hand sides of (9) and (10) can now be termed the marginal product of L in the production of z of a given firm. Let W be the price of the production factor L. By substituting in the RHS of equation (18′) in S-B (p. 86), $W/MP_L(X)$ for $MC(X)$, and then multiplying the resulting RHS and the middle term in the chain (18′) there by $MP_L(X)$ we obtain

$$(11)\quad P_x MP_L(X)(1+1/\varepsilon_{\pi_z}) = W,$$

where ε_{π_z} is the demand elasticity of z and x.

In much the same way we substitute $W/MP_L(A)$ for P_A in (19′) in (S-B) and multiply both sides of the equation by $MP_L(A)$ to obtain

$$(12)\quad \pi_z g' MP_L(A) X(1+1/\varepsilon_{\pi_z}) = W,$$

since the LHS of both (11) and (12) equal W they also equal each other, hence

$$(13)\quad P_x MP_L(X)(1+1/\varepsilon_{\pi_z}) = \pi_z g' X(1+1/\varepsilon_{\pi_z}) MP_L(A).$$

By utilizing $\pi_z = P_x/g(A)$, eliminating terms from both sides of (11) and rearranging terms we get equation (10) of the text above.

The fact that equation (10) follows from equation (18′) and (19′) in S-B for all values of ε_{π_z} implies first that the allocation of resources by the firm to x and A in the production of z is efficient whether the firm operates in competitive or monopolized markets. Second, that the quantity of advertising used in the production of z is determined solely by production conditions and is not effected by demand elasticity beyond the effect of demand on the quantity of z produced. Thus a given quantity of z will be produced by given quantities of A and x regardless of demand elasticity.[2]

We have shown above that condition (10), being a production efficiency condition within the firm, holds under any market arrangement as long as the individual firm is maximizing profits.

Another necessary efficiency condition is one requiring equality between the rate of production transformation (RPT) of the commodity z and the good y, and the rate of substitution in consumption (RCS) of the same two goods. In equational form this condition is as follows.

$$(14)\quad \mathrm{RCS} = \frac{\partial u(z,Y)}{\partial z} \Big/ \frac{\partial u(z,y)}{\partial y} = \frac{MP_L(y)}{g(A)MP_L(X)} = \mathrm{RPT}.$$

It should be noted that the above relation is restricted to customers and production of a given brand of z. The rationale for relation

---

[2] In their paper Stigler and Becker comment that "...the optimal level of advertising would be positively related to the commodity elasticity" (p. 86). Obviously, demand elasticity has no direct effect on advertising and it only effects advertising through its effect on the quantity of Z produced.

By equating the RHS of equation (7) in Nichols to zero, and performing proper substitutions, we again obtain equation (10) of the above text. Thus Nichols' interpretation of his equation (7) and its efficiency implications are not valid.

(14) is that, if (14) does not hold but rather an inequality holds in (14), we could by shifting factors from the production of z to the production of y (or vice versa) substitute z with y through production so, that the added gain in utility due to the additional quantity of the increased good exceeds the loss in utility due to the reduction in quantity of the second good.[3]

Returning to the market allocation, since $\partial u/\partial x = (\partial u/\partial z)\cdot g(A)$, from (7) we obtain

$$\frac{\partial u/\partial x}{\partial u/\partial y} = \frac{(\partial u/\partial z)g(A)}{\partial u/\partial y} = \frac{P_x}{P_y};$$

hence

$$\frac{\partial u/\partial z}{\partial u/\partial y} = \frac{P_x/g(A)}{P_y} = \frac{\pi_z}{P_y}. \tag{15}$$

Since y is produced by a competitive industry the following holds,

$$P_y MP_L(y) = W. \tag{16}$$

By substituting, for W in (16), the LHS of (11), then substituting $\pi_z\cdot g(A)$ for P_x, and rearranging terms we obtain

$$\frac{\pi_z}{p_y} = \frac{MP_L(y)}{g(A)MP_L(x)(1+1/\varepsilon_{\pi_z})}. \tag{17}$$

By substituting (17) in (15), we obtain that in the market the following holds,

$$\frac{MP_L(y)}{g(A)MP_L(x)(1+1/\varepsilon_{\pi_z})} = \frac{(\partial u/\partial z)}{(\partial u/\partial y)}. \tag{18}$$

Upon comparing (18) with the efficiency condition (14) we learn that the market solution fulfills this efficiency requirement if and only if,[4]

$$\varepsilon_{\pi_z} \equiv \varepsilon_{p_x} = -\infty, \tag{19}$$

which is a necessary condition for the efficiency of a market allocation.[5]

Thus, as in traditional economic theory, efficiency is attained only if producers are product price takers, and this time, commodity price takers as well.

A situation in which customers are linked to a brand and cannot change it costlessly may occur for example, when different brands are produced in different geographical areas, or when changing brands involves a costly adjustment process of facilities, knowledge, habits, etc. In these situations, as in classical microeconomic theory, efficiency is unlikely to occur since demand elasticity facing producers is finite, and (19) is not fulfilled.

In many cases, however, changing brands by customers is easy and involves little or no cost, like changing the brands of a soft drink, changing the make of a car, or changing an airline, etc. When individuals can move freely from one brand of z to another, the price of z facing producers is fixed, and the demand elasticity facing each producer is $-\infty$. Furthermore, in this case all producers are facing the same commodity price π_z.

It should be noted that the same commodity price to two firms does not necessarily imply the same product price. The proof and implications of this fact are discussed fully below.

The assumption of free mobility of customers from one brand to another leads us to the inter-firm, intra-industry efficiency condition

$$MP_L(Z_i) = MP_L(Z_k), \tag{20}$$

keeping in mind, that from (10), $MP_L(Z_i)$ is uniquely defined for each brand i and that i,k, $i \neq k$ vary over all possible brands of the commodity z, each produced by a different firm.

---

[3] This argument is identical to the accepted argument concerning the necessity for efficiency, of having equality between RCS and RPT. This can be found in any welfare economics textbook. Therefore, no further elaboration on this topic is offered.

[4] See S-B, p. 86, fn. 15, for proof that $\varepsilon_{\pi_z} = \varepsilon_{p_x}$.

[5] This condition has not been observed by Nichols to be consistent with efficiency. Instead he states that efficiency is attained when the market for x is not perfectly competitive, that is, $\varepsilon_{p_x} > -\infty$ (E_{x,p_x} in Nichols' notation, see bottom p. 216, ibid., Case I)

The rationale of (20) is as follows: Suppose that, instead of equality, inequality holds in the above equation (20). Then we could shift a customer from the brand with low-marginal product to a brand with high-marginal product and shift production factors in the same direction so that the production of the commodity by the low-marginal product firm should decline by exactly the amount consumed by the shifted customer. Then the high-marginal product firm will produce, by using the added production factors, additional quantity of its brand of the commodity, which is, because of its higher MP_L, larger than the quantity reduced from the output of the low-marginal product firm. Therefore, the utility of the shifted customer will increase due to the shift, while the utility of all other customers in the economy remain unchanged. This implies that an allocation in which inequality holds in (20) can be improved and hence is not optimal. This, in turn, implies that equality in (20) is a necessary condition for Pareto efficiency.

For condition (20) to hold in a market allocation what is needed is for the firm to be a commodity price taker, (which is also required for the fulfillment of (14)), and for the commodity price π_z to be fixed and equal to all firms. This is proved by substituting in equation (17) $\varepsilon_{\pi_z} = -\infty$ and noting that π_z is the same for all z-producing firms, equation (20) now follows immediately from (17). This in turn implies $g(A_i)MP_L(X_i) = g(A_k)MP_L(X_i)$, for all $i \neq k$, which in turn implies (20).

The existence of many firms in the economy, each producing the same commodity and competing among themselves for the same customers indeed guarantees the same commodity price to all firms, and by it guarantees efficiency.

In summary in this section we have derived a full set of independent efficiency conditions[6] of S-B's economy, this being equations (2), (10), (14), (20). These conditions are fulfilled if and only if the economy, including the commodity market, is perfectly competitive, that is, in the advertised industry all firms are commodity price takers. When firms are facing finite demand elasticities in some of the good and/or in commodity markets, efficiency is not attained.[7]

II. Characterization of Solutions

First, we consider two firms of our advertised industry, operating in the same competitive market, having different production functions and therefore each producing different output levels. Both firms are facing the same commodity price, still the larger of the two firms (the one producing more x) will provide more advertising to its customers, a unit of its commodity will consist of more advertising and less of the product and its product price will be higher.

To prove the above assertion, let F_1 and F_2 be the two production functions of the two above-mentioned firms. Without loss of

---

[6] The necessary efficiency conditions discussed and proved as such in this section, also constitute and exhaust a mutually independent set of necessary conditions to the general optimization problem of S-B's economy. Thus they constitute a full set of first-best conditions.

[7] Nichols' Case I (ibid., p. 216) cannot exist. On one hand he assumes that ε_{p_x} (E_{x,p_x} in his notation) is finite and on the other hand he assumes that the demand price π_z (P_z in his notation) fulfills $\partial\pi_z/\partial A = 0$. These two assumptions contradict each other. To prove that, we carry out the following derivation.

$$\partial\pi_z/\partial A = (d\pi_z/dz)(\partial z/\partial A) = (d\pi_z/dz)g'(A)x$$

$$= (\pi_z g'(A)x)/(z\varepsilon_{\pi_z})$$

$$= [P_x g'(A)]/\left[\varepsilon_{p_x}(g(A))^2\right].$$

It should be noted that $\varepsilon_{p_x} = \varepsilon_{\pi_z}$; see fn. 4. Since P_x, $g(A)$ and $g'(A)$ are all finite, it follows that $\partial\pi_z/\partial A$ can vanish if and only if ε_{p_x} is infinite. Thus Nichols' Cases I and II cannot exist. Therefore, out of the four market situations Nichols mentions only two exist. The first is the one in which both $\varepsilon_{\pi_z} \equiv \varepsilon_{p_x} = -\infty$ and the second is one in which $|\varepsilon_{p_x}| \equiv |\varepsilon_{\pi_z}| < \infty$. Indeed, these are the two cases investigated by S-B and in our paper. Thus, only Cases III and IV of Nichols really exist, yet he does not study them and instead concentrates on the nonexisting cases.

generality, assume that the marginal product of firm 1 is larger than that of firm 2, when both are producing the same quantity of X; that is, $MP_L^1(X) > MP_L^2(X)$, for all positive X.

Now consider equations (10) and (20), both of which are fulfilled in perfect competition. Let $MP_L(Z)$ be the common marginal product of the commodity Z, then,

$$(21)\quad \begin{aligned} MP_L(Z) &= g(A_1)MP_L^1(X_1) \\ &= g(A_2)MP_L^2(X_2) \\ &= MP_L(A)g'(A_1)X_1 \\ &= MP_L(A)g'(A_2)X_2, \end{aligned}$$

where A_i and X_i are respectively the advertising and product of firm i.

The solution to (21) must fulfill the conditions

$$(22)\qquad X_1 < X_2 \text{ and } A_1 < A_2,$$

any other relations will lead to a contradiction. For example, suppose $X_1 = X_2$, then the last equality in the chain above implies that $A_1 = A_2$. This in turn implies that the second equality in the chain above cannot hold since we assumed $MP_L^1(X_1 = X_2) > MP_L^2(x_2)$ while $g(A_1) = g(A_2)$, a contradiction.

Thus, if we assume[8] $F'' < 0$ and $g'' < 0$, the two following relations must hold $g'(A_1) > g'(A_2)$ and $MP_L^1(X_1) > MP_L^2(X_2)$, which leads to (22). Thus a larger firm is producing more output as well as using more advertising. Therefore a unit of the commodity z produced by the larger firm contains less of the product and more advertising than a unit of commodity produced by a smaller firm (the amount of x per unit z for firm j is $X_j/Z_j = 1/g(A_j)$). The market price, P_x, of the product of the larger firm is higher as well. To see that, note that $P_x = \pi_z \cdot g(A)$, and since π_z is the same for all firms and $g(A_1) < g(A_2)$ the result is straightforward.

Second, we consider our industry with different concentrations. We argue that the amount of concentration in a competitive industry also effects the ratios of X and A in the production of Z. An industry with high concentration has few large firms instead of many small ones in an industry with low concentration. An increase in a firm's size implies both an increase in advertising by the firm and in the output level of the firm.[9] This in turn leads to an increase in product prices and a decrease in quantity of product per unit commodity as shown above.

All of the above results characterize firms facing finite demand elasticities as well. The only certain way, therefore, to distinguish between a perfectly competitive industry and one engaged in monopolistic competition is to find out whether in each firm in the industry product price equals marginal cost of production. In the case of monopolistic competition this price is above marginal cost, and in perfectly competitive commodity market, product price equals marginal cost of production, even though both the product price and marginal cost may vary from one firm to another.

Since Chamberlin and Robinson, a market in which each firm sells its product at a difference price and uses advertising is believed to be engaged in monopolistic competition and therefore to operate inefficiently. Stigler and Becker (1977) provide us with a

---

[8] If any or both assumptions did not hold, then we would be in the realm of economies of scale, in which case efficiency and market conditions lead to a single-operating firm in the industry. The market allocation is a single-monopolistic firm which operates inefficiently.

[9] We assume constant marginal product in advertising when factor price does not change (due to RCS in the advertising industry). Then differentiating (10) totally and rearranging terms we obtain

$$dA/dX = \frac{MP_L(A)\cdot g' - g\cdot d(MP_L(X))/dX}{g'\cdot MP_L(X) - MP_L(A)\cdot g''\cdot X} > 0,$$

where g is $g(A)$ and $g' = dg(A)/dA$ and $g'' = d^2g(A)/dA^2$. The inequality in the above equation follows when diminishing marginal product is assumed in the production of x, $[d(MP_L(X))/dX < 0]$ and in the household production of z with respect to A, $(g''(A) < 0)$.

model which utilizes the "new theory of consumption," advanced by both Becker and Kelvin Lancaster (1971) in the 1960s and 1970s to bring us back to the realm where the invisible hand of the market operates efficiently.

REFERENCES

Chamberlin, Edward, H., *The Theory of Monopolistic Competition*, Cambridge: Harvard University Press, 1933.

Lancaster, Kelvin, *Consumer Demand: A New Approach*, New York and London: Columbia University Press, 1971.

Michael, R. T. and Becker, Gary, S., "On the New Theory of Consumer Behavior," *Swedish Journal of Economics*, December 1973, *75*, 378–96.

Nichols, Len, M., "Advertising and Economic Welfare," *American Economic Review*, March 1985, *75*, 213–18.

Robinson, Joan, *Economics of Imperfect Competition*, London-New York: Macmillan and St. Martin's Press, 1965.

Stigler, George J. and Becker, Gary S., "De Gustibus Non Est Disputandum," *American Economic Review*, March 1977, *67*, 76–90.

Diamonds Are a Government's Best Friend: Burden-Free Taxes on Goods Valued for Their Values: Comment

By DAVID D. FRIEDMAN*

In a recent article (Yew-Kwang Ng, 1987), Ng demonstrates that taxes on goods whose utility is derived entirely from their market value ("diamond goods") can provide revenue to the government with no cost to the taxpayers aside from the administrative costs of collecting them. Such a tax decreases the quantity of the good produced while proportionally increasing the value per unit quantity. Since the utility of the good depends only on its value, the smaller quantity produced (after the tax is imposed) produces the same total utility to consumers as the previous larger quantity. The revenue collected by the government is paid, in effect, from the reduced cost of producing a smaller quantity of the good.

Ng mentions a number of earlier discussions of goods whose utility comes in part from their price. He is apparently unaware of an analysis which is both much earlier and much closer to his than any he cites. In Chapter 13 of the *Principles of Political Economy*, David Ricardo wrote:

> If gold were the produce of one country only, and it were used universally for money, a very considerable tax might be imposed on it, which would not fall on any country, except in proportion as they used it in manufactures, and for utensils; upon that portion which was used for money, though a large tax might be received, nobody would pay it. ...The benefit would be this, that if less gold were produced, less capital would be employed in producing it; ...From such a tax, as far as money was concerned, the nations of Europe would suffer no injury whatever; they would have the same quantity of goods, and consequently the same means of enjoyment as before, but these goods would be circulated with a less quantity, because a more valuable money. ...If in consequence of the tax, only one tenth of the present quantity of gold were obtained from the mines, that tenth would be of equal value with the ten tenths now produced. [Ricardo, 1951, pp. 194–96]

As can be seen from the quotation, Ricardo's burden-free tax is a tax on gold used exclusively for money. A reduction in the quantity of gold results in a proportional increase in its price; since the services provided by a coin depend on what it will buy, not how much it weighs, the smaller quantity produced as a result of the tax provides the same monetary services as the larger quantity produced before the tax. The revenue to the government is paid by the reduced cost of producing a smaller quantity of gold.[1]

Ng's analysis is more general than Ricardo's, and considerably clearer. The essential argument, however, is exactly the same.

REFERENCES

Ng, Yew-Kwang, "Diamonds Are a Government's Best Friend: Burden-Free Taxes on Goods Valued for Their Values," *American Economic Review*, March 1987, *77*, 186–91.

Ricardo, David, *The Principles of Political Economy*, Piero Sraffa, ed., New York: Cambridge University Press, 1951.

---

*Faculty Fellow in Law and Economics, University of Chicago Law School, 1111 E. 60th St., Chicago, IL 60637.

[1] One minor difference between the cases by Ng and by Ricardo is that that Ng assumes a horizontal supply curve, while Ricardo assumes that there exist mines of differing quality. The result is that in Ricardo's example government revenue represents in part a savings in production cost and in part a transfer from the mine owners to the government.

The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67: A Critique and Update of Wolff's Estimates

By FRED MOSELEY*

The most important prediction of Karl Marx's theory of capitalism is that the rate of profit would tend to decline over time as a result of technological change. According to Marx's theory, the rate of profit varies directly with the rate of surplus value and inversely with the composition of capital. Marx argued that both the composition of capital and the rate of surplus value would increase as a result of technological change, thus having offsetting effects on the rate of profit. However, Marx argued further that the composition of capital would increase at a faster rate than the rate of surplus value, so that the net effect would be a decline in the rate of profit. (Marx, 1977, Pts. 4 and 7; and Marx, 1982, Pt. 3).[1]

In an important paper in Marxian empirical research in this *Review*, Edward Wolff (1979) presented estimates of these three key variables in Marx's theory for the U.S. economy from 1947 to 1967. According to Wolff's estimates, both the composition of capital and the rate of surplus value increased, as predicted by Marx's theory; however, the rate of surplus value increased at a faster rate than the composition of capital, so that the rate of profit increased, contrary to Marx's prediction. Wolff concluded that "there is no empirical support for it (Marx's law of the tendency of the rate of profit to fall) during my period of investigation" (p. 340).

In a more recent paper, Wolff (1986) revised and updated these estimates to 1976 and came to somewhat different conclusions. According to his estimates for the entire 1947–76 period, the rate of profit declined, as predicted by Marx's theory, but this decline was due to a decline in the rate of surplus value rather than to an increase in the composition of capital, contrary to Marx's prediction.[2]

I argue that Wolff's estimates, although clearly an important contribution to Marxian empirical research, do not provide a reliable empirical test of Marx's theory because these estimates are derived from data categories which differ in significant respects from the basic concepts in Marx's theory.[3] The most

---

*Department of Economics, Colby College, Waterville ME 04901. I wish to express a special thanks to Ed Wolff for providing me with his data, for answering many detailed and time-consuming questions, and for penetrating comments on an earlier draft of this paper. I have also benefited from comments by Tom Weisskopf, James Crotty, Mark Glick, Ahmed Tonak, and anonymous referees of this journal. I also thank Randy Mitchell, Kariena Zarcharski, and William Ralph for their very capable research assistance and Colby College for financial assistance.

[1] There has been much theoretical controversy over whether Marx conclusively proved that the rate of profit must fall as a result of technological change. See Jens Christiansen for a concise review of these controversies. This paper does not enter into this long and continuing debate. Instead, this paper is concerned solely with the empirical question of whether the actual trends of the three Marxian variables were in the directions which Marx at least expected and asserted, if not conclusively proved.

[2] Another important study which also presents estimates of the three key Marxian variables for the postwar U.S. economy is Thomas Weisskopf, which comes to essentially the same conclusion as Wolff: The rate of profit declined (25 percent) because the rate of surplus value declined (28 percent), with essentially no trend in the composition of capital.

[3] The problem of the relation between theoretical concepts and the data used to estimate these concepts is of course not unique to Marxian empirical research. A recent example of a similar controversy within this *Review* related to the neoclassical theory of monopoly is whether the accounting rate of profit is a reliable proxy for the economic rate of profit, which according to neoclassical theory is the appropriate indicator of monopoly profits within an industry. The controversy was initiated by Franklin Fisher and John McGowan.

important discrepancy between the Marxian concepts and Wolff's estimates is that the latter do not take into account Marx's distinction between productive capital and unproductive capital. Section I below briefly discusses this important distinction in Marx's theory. Section II presents estimates of the Marxian variables which take this distinction into account. Section III compares these estimates with Wolff's estimates of the Marxian variables.

I. Productive Capital and Unproductive Capital[4]

The three Marxian variables discussed above are ratios defined in terms of three fundamental concepts: constant capital, variable capital, and surplus value. The composition of capital (CC) is defined as the ratio of the accumulated stock of constant capital (C) to the annual flow of variable capital (V); that is $CC = C/V$. The rate of surplus value (RS) is defined as the ratio of the annual flow of surplus value (S) to the annual flow of variable capital; that is, $RS = S/V$. Finally, the rate of profit is defined as the ratio of the annual flow of surplus value to the sum of the accumulated stock of constant capital; that is, $RP = S/C$. (I make here the convenient simplification that the stock of variable capital is zero, since it is in actuality negligibly small.)

If both terms of the rate of profit are divided by the flow of variable capital, we obtain the following accounting relationship, which provides the analytical framework for Marx's theory of the rate of profit,

$$(1) \qquad RP = \frac{S/V}{C/V} = \frac{RS}{CC}.$$

Thus it is clear that, according to this framework, the rate of profit varies directly with the rate of surplus value and inversely with the composition of capital, as stated above.

The most important conceptual issue involved in the precise definition of these fundamental concepts and thus in the estimation of the three Marxian ratios is whether the concepts of constant capital and variable capital include all the capital invested in capitalist enterprises, or only the capital invested in production activities.[5] I argue that Marx's concepts of constant capital and variable capital include only the capital invested in *production activities*, where "production" is defined fairly broadly to include such activities as transportation and storage, but specifically does not include the following two types of activities within capitalist enterprises: 1) *Circulation activities*, or activities related to the buying and selling of commodities, including such activities as sales, purchasing, accounting, advertising, debt-credit relations, legal counsel, etc.; and 2) *Supervision activities*, or activities related to the control of the labor of production workers, including such activities as management, supervision, record keeping, etc.

This distinction between production and nonproduction activities is based on Marx's theory of value and surplus value, according to which value and surplus value are created in production activities, but are not created in nonproduction activities (Marx, 1977, chs. 1, 7, and 13; Marx, 1981, ch. 6; and Marx, 1982, ch. 17). Thus Marx referred to capital invested in production activities as "productive capital" and capital invested in nonproduction activities as "unproductive capital."

I argue that constant capital and variable capital were defined by Marx as components of "productive capital." Constant capital is the capital invested in the means of *production*; it does not include the capital invested in the means of circulation or in the means of supervision. Similarly, variable capital is the capital invested in labor power utilized in production; it does not include the capital invested in labor power employed in circulation and supervision activities.[6]

---

[4] This section is a brief summary of a longer discussion in Moseley (1986b).

[5] Other conceptual issues involved in the precise definition of these Marxian concepts and thus in the estimation of the Marxian ratios include: 1) whether the Marxian concepts should be estimated in terms of labor-hours or in terms of prices; 2) whether the concepts of constant capital and variable capital refer only to capitalist production or also refer to various forms of noncapitalist production (mainly government production, but also household and institutional production); and 3) whether the taxes on wages are a part of variable capital or surplus value. These issues are discussed in Fred Moseley (1986b). Briefly stated, I argue that 1) the Marxian concepts should be estimated in terms of prices, 2) the concepts of constant capital and variable capital refer only to capitalist production, and 3) the taxes on wages are a part of variable capital. Wolff adopts the opposite interpretation of all these issues.

Wolff does not take into account Marx's distinction between productive capital and unproductive capital in his estimates of the Marxian variables. Instead, Wolff defines constant capital as the current value of *all* structures, equipment, and materials utilized within capitalist enterprises, including those utilized in the functions of circulation and supervision. Similarly, he defines variable capital as the wages of *all* labor within capitalist enterprises, including the labor employed in circulation and supervision.[7]

II. Estimates of the Marxian Variables

Annual estimates of the Marxian variables of constant capital, variable capital, and surplus value were derived for the U.S. economy over Wolff's period of 1947–76, based on the definitions given in Section I. The data sources and methods used to derive these estimates are described briefly in the Appendix and more fully in the references cited therein.

TABLE 1—ESTIMATES OF THE MARXIAN VARIABLES

	Composition of Capital	Rate of Surplus Value	Rate of Profit
1947	3.46	1.40	0.40
1948	3.48	1.35	0.39
1949	3.68	1.50	0.41
1950	3.81	1.42	0.37
1951	3.66	1.44	0.39
1952	3.58	1.41	0.39
1953	3.46	1.35	0.39
1954	3.73	1.46	0.39
1955	3.74	1.51	0.40
1956	3.84	1.44	0.37
1957	3.95	1.50	0.38
1958	4.19	1.59	0.38
1959	4.00	1.61	0.40
1960	3.97	1.62	0.41
1961	4.03	1.68	0.42
1962	3.91	1.71	0.44
1963	3.83	1.71	0.45
1964	3.76	1.73	0.46
1965	3.75	1.73	0.46
1966	3.74	1.72	0.46
1967	3.86	1.72	0.45
1968	3.84	1.69	0.44
1969	3.88	1.62	0.42
1970	4.08	1.61	0.40
1971	4.27	1.71	0.40
1972	4.13	1.67	0.40
1973	4.15	1.59	0.38
1974	4.86	1.55	0.32
1975	5.11	1.71	0.33
1976	4.88	1.66	0.34

Sources: See the Appendix.

From these estimates of the basic magnitudes, estimates of the three key Marxian ratios were calculated. These estimates are presented in Table 1. These estimates show the following trends: The estimates of the composition of capital increased 41 percent over this period, from 3.46 in 1947 to 4.88 in 1976, as predicted by Marx's theory. Measured in terms of five-year averages, the composition of capital increased 28 percent, from an average of 3.62 in 1947–51 to an average of 4.63 in 1972–76.

The estimates of the rate of surplus value increased 19 percent over this period, from 1.40 in 1947 to 1.66 in 1976, again as predicted by Marx's theory. In terms of five-year averages, the rate of surplus value increased

---

[6] Important references on Marx's concepts of productive capital and unproductive capital include I. Rubin, 1972; Ian Gough, 1972; Harry Braverman, 1974; and David Leadbeater, 1985.

[7] Wolff states in a footnote in his 1986 paper that he ignores the distinction between productive labor and unproductive labor because his primary aim is to explain trends in the "capitalist" rate of profit, not the Marxian rate of profit. However, his 1979 paper is explicitly presented as an empirical test of Marx's theory, and his 1986 paper is essentially an update of his 1979 paper using the same Marxian variables and the same estimation procedures. Thus the reader is certainly left with the impression that the 1986 paper is intended, at least in part, to be an empirical test of Marx's theory. Beyond this question, even if Wolff's primary aim in the 1986 paper is to explain the trends in the "capitalist" rate of profit, that in itself is no reason to ignore Marx's distinction between productive labor and unproductive labor. Moseley (1986b) presents an analysis of the "capitalist" rate of profit, in which this distinction plays a crucial role.

15 percent, from 1.42 in 1947–51 to 1.64 in 1972–76.

Finally, since the increase in the composition of capital was proportionally greater than the increase in the rate of surplus value, the rate of profit fell significantly during this period. The estimates of the rate of profit declined 15 percent, from 0.40 in 1947 to 0.34 in 1976. In terms of five-year averages, the rate of profit declined 10 percent, from 0.39 in 1947–51 to 0.35 in 1972–76. This decline in the rate of profit is of course consistent with the prediction of Marx's theory.

III. Comparison with Wolff's Estimates

The estimates of the Marxian ratios presented in Section II will now be compared with Wolff's estimates of these ratios. The main differences between these two sets of estimates is that the estimates of the composition of capital and the rate of surplus value have essentially opposite trends. Wolff's estimates of the composition of capital declined 8 percent over his period (from 5.62 in 1947 to 5.16 in 1976), while my estimates of the composition of capital increased 41 percent, as we have just seen. Both estimates are shown graphically in Figure 1 (my estimates are shown in terms of five-year moving averages and both estimates are converted into index numbers to emphasize the secular trends).[8]

Similarly, Wolff's estimates of the rate of surplus value declined 27 percent over this period (from 0.96 in 1947 to 0.75 in 1976), while my estimates of the rate of surplus value increased 19 percent. These two estimates of the rate of surplus value are shown in Figure 2 (converted as above).[9]

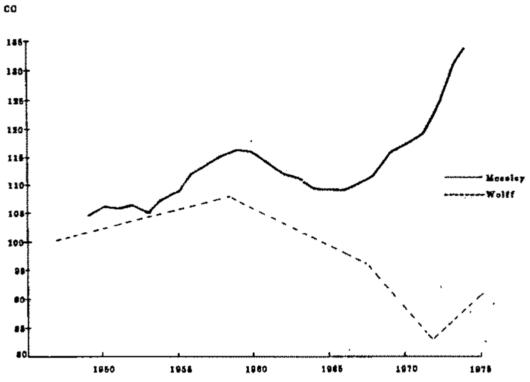


FIGURE 1. ESTIMATES OF THE COMPOSITION OF CAPITAL

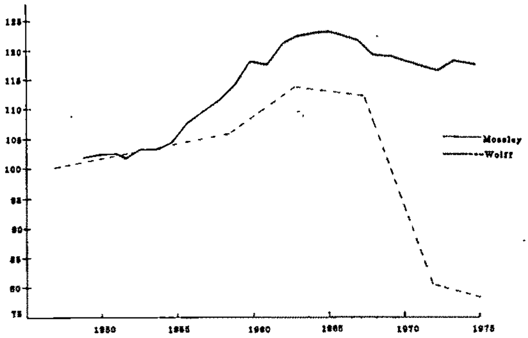


FIGURE 2. ESTIMATES OF THE RATE OF SURPLUS VALUE

These two sets of estimates of the Marxian ratios suggest entirely different causes of the decline in the rate of profit in the postwar U.S. economy.[10] As noted in the introduction, Wolff's estimates suggest that the decline in the rate of profit was caused by a decline in the rate of surplus value, contrary to Marx's predictions. By contrast, my estimates suggest that the decline in the rate of profit was caused by a greater increase in the composition of capital than in the rate of profit, as Marx predicted.

---

[8] Wolff's estimation methodology, unlike my methodology, involves the use of input-output tables, which are available for the postwar U.S. economy only at irregular intervals of 5–10 years, so Wolff's time-series for each variable consists of only 6 estimates.

[9] These divergent trends in the composition of capital and the rate of surplus value had roughly offsetting effects on the trend in the rate of profit, so that both estimates of the rate of profit show a declining trend (Wolff's estimates declined 20 percent compared to the 10–15 percent decline in my estimates).

[10] "Cause" is used here in the sense of a proximate cause. An analysis of the underlying determinants of the composition of capital and the rate of surplus value is beyond the scope of this discussion.

These divergent trends in the composition of capital and the rate of surplus value are almost entirely explained by the different treatments given to Marx's distinction between productive capital and unproductive capital by Wolff and myself. This conceptual difference accounts for 44 percent of the difference in the trends of the two estimates of the composition of capital and for 85 percent of the difference in the trends of the two estimates of the rate of surplus value.[11]

Thus the conclusion one reaches concerning the trends of the composition of capital and the rate of surplus value and hence concerning the cause of the decline in the Marxian rate of profit in the postwar U.S. economy depends almost entirely on one's interpretation of Marx's distinction between productive capital and unproductive capital.[12] If the distinction is taken into account, as I have argued it should be in an empirical test of Marx's theory, then the decline in the rate of profit in the postwar U.S. economy was caused by an increase in the composition of capital, not by a decline in the rate of surplus value.

APPENDIX: DATA SOURCES AND METHODS

1. *Constant capital* (*stock*): the sum of fixed capital and circulating capital. *Fixed capital*: the current value of buildings and equipment used for production activities. Estimates are derived from the Bureau of Economic Analysis (BEA) data for "net private nonresidential fixed capital" (current cost), excluding various types of buildings and equipment used for circulation and supervision activities (for example, office, computing, and accounting machines; furniture and fixtures; commercial buildings; etc.). *Circulating capital*: the current value of inventories. Estimates are derived from NIPA data for "business inventories." Detailed description in Moseley (1985b).

2. *Variable capital* (*annual flow*): the total compensation (including supplements and benefits) of production workers. Estimates are derived from NIPA data for "total employee compensation" in the business sector of the economy, excluding the compensation of nonproduction workers. The percentage of total employee compensation within each of the eight major industry classifications that was paid to production workers is estimated using data from various sources, primarily the Censuses of Manufacturing, Mining, and Construction, which provides data for the wages of "production workers," and the U.S. Department of Labor, Bureau of Labor Statistics (BLS), *Employment and Earnings*, (*Establishment Survey*), which provides data for the numbers of "production workers" in the manufacturing, mining, and construction industries and for the number of "nonsupervisory employees" in the other industries. Detailed description in Moseley (1985a).

3. *Surplus Value* (*annual flow*): the difference between new value and variable capital. Estimates of *new value* (or "value added") are derived from NIPA data for the "net product" of the business sector, excluding "imputations" which do not correspond to goods and services actually sold on the market (80 percent of which is the value of

---

[11] Most of the remainder of the difference in the trends of the two estimates of the composition of capital is accounted for by the different treatments of noncapitalist production (see fn. 4). The other two conceptual issues discussed in fn. 4 had very little effect on the trends of the Marxian ratios in the postwar U.S. economy. The method used to estimate the contribution of each of these conceptual issues to the difference in the trends of my estimates and Wolff's estimates is adopted from Edward Denison (1960), which reconciles Denison's estimates of inputs and output in the U.S. economy (1950–62) with Jorgenson and Griliches' estimates of these variable.

[12] Essentially the same conclusion is reached by a similar comparison of my estimates and Weisskopf's estimates of the Marxian ratios. It was noted above (fn. 2) that the trends in Weisskopf's estimates are very close to the trends in Wolff's estimates. The differences between the trends of my estimates and Weisskopf's estimates are also due primarily to the fact that Weisskopf does not take into account Marx's distinction between productive capital and unproductive capital. This conceptual difference accounts for 68 percent of the difference in the two trends of the composition of capital and for 65 percent of the difference in the two trends at the rate of surplus value. For a critique of Weisskopf's estimates of the rate of surplus value, see Moseley (1985a).

the "housing services" of owner-occupied homes). Detailed description in Moseley (1985a).

REFERENCES

Braverman, Harry, *Labor and Monopoly Capital*, New York: Monthly Review Press, 1974.

Christiansen, Jens, "Marx and the Falling Rate of Profit," *American Economic Review*, May 1976, *66*, 20–26.

Denison, Edward F., "Some Major Issues in Productivity Analysis: An Examination of Estimates by Jorgenson and Griliches," *Survey of Current Business*, May 1969, Pt. II, *49*, 1–27.

Fisher, Franklin M. and McGowan, John J., "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits," *American Economic Review*, March 1983, *73*, 82–97.

Gough, Ian, "Marx's Theory of Productive and Unproductive Labor," *New Left Review*, November-December 1972, 47–72.

Leadbeater, David, "The Consistency of Marx's Categories of Productive Labor and Unproductive Labor," *History of Political Economy*, Winter 1985, *17*, 591–619.

Marx, Karl, *Capital*, Vol. 1, New York: Random House, 1977.

______, *Capital*, Vol. 2, New York: Random House, 1981.

______, *Capital*, Vol. 3, New York: Random House, 1982.

______, *Theories of Surplus Value*, Moscow: Progress Publishers, 1968.

Moseley, Fred, (1985a) "The Rate of Surplus Value in the Postwar U.S. Economy: A Critique of Weisskopf's Estimates," *Cambridge Journal of Economics*, March 1985, *9*, 57–79.

______, (1985b) "Estimates of the Composition of Capital in the Postwar U.S. Economy," presented at the 1985 Allied Social Sciences Association Annual Convention.

______, (1986a) "Estimates of the Rate of Surplus Value in the Postwar U.S. Economy," *Review of Radical Political Economics*, Spring and Summer 1986, *18*, 168–89.

______, (1986b) "Marxian Empirical Research and the Postwar U.S. Economy: A Critique of Wolff's Estimates," presented at the 1986 Allied Social Sciences Association Annual Convention.

Rubin, I. I., *Essays in Marx's Theory of Value*, Detroit: Black and Red Press, 1972.

Weisskopf, Thomas E., "Marxian Crisis Theory and the Rate of Profit in the Postwar U.S. Economy," *Cambridge Journal of Economics*, December 1979, *3*, 341–78.

Wolff, Edward, "The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67," *American Economic Review*, June 1979, *69*, 329–41.

______, "The Productivity Slowdown and the Fall in the Rate of Profit, 1947–76," *Review of Radical Political Economy*, Spring-Summer 1986, *18*, 87–109.

The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67: Reply

By EDWARD N. WOLFF*

Fred Moseley's paper is a very useful contribution to an important topic in Marxian economics, factors which affect movements in the rate of profit over time. Its main contribution is to include consideration of unproductive activity in the definition and movement of the rate of profit. The value added by unproductive workers is included as part of surplus value, whereas the stock of capital invested in unproductive activity is excluded from constant capital. Thus, the "Marxian" rate of profit, defined as the ratio of surplus value to constant capital, will be greater than the conventional or "capitalist" rate of profit.

As shown in Moseley's paper, the rate of profit *RP* is the ratio of the rate of surplus value *RS* to the (organic) composition of capital *CC*. *RS* is defined as the ratio of surplus value *S* to variable capital *V* and the organic composition of capital as the ratio of constant capital *C* to *V*. The issue of whether to include unproductive capital and labor in the definition of the organic composition is a debatable one. The argument for its inclusion is that from the capitalist's point of view, payment for unproductive labor constitutes a cost just as much as wages paid to productive labor. Forces of competition necessitate that the capitalist obtain the average rate of return on the capital advanced to unproductive labor. Moreover, unproductive constant capital, the inputs, and capital stock employed in unproductive activity should likewise be included in the definition of the organic composition, since the capitalist expects the normal rate of return on this capital. Thus, from the point of view of the capitalist rate of profit, the wages paid to unproductive labor and the expenditures on unproductive capital should enter cost, not surplus value, and thus unproductive wages should be included in variable capital and unproductive capital in constant capital. The argument against the inclusion of unproductive inputs in the calculation of the organic composition is that they are paid for out of surplus value and thus do not, properly speaking, constitute capital. This is Moseley's position, and it is also quite legitimate.

There are four other differences between Moseley's estimates and mine. First, Moseley excludes the capital of the government sector from the definition of constant capital and the labor employed by the government from variable capital on the grounds that government activity is unproductive. Yet, the government sector provides many goods and services that are comparable to those provided by productive private enterprises. For example, the government sector includes what are called "government enterprises," which are businesses that operate in almost all respects like capitalist enterprises and whose product or services are sold directly to customers. Moreover, many governmental activities which are financed by tax dollars provide services directly to the public. These include public educational systems, local water and sewerage systems, garbage collection, fire protection, highways, and the like. In many cases, there are close substitutes provided by capitalist enterprises, such as garbage collection and public transit in many localities.

Thus, many governmental activities are indistinguishable from those of the private sector. This suggests that government capital and labor be incorporated in the organic composition. The argument against this is that there is no reason for the government to try to obtain the average (capitalist) rate of return on the capital it advances. Indeed, the forces of competition, which cause the rate of profit to be equalized in the private sector, do not operate in the government sector. As a result, it appears that both arguments have legitimacy and both treatments of government capital are acceptable.[1]

---

*Department of Economics, New York University, New York, NY 10003.

Second, Moseley uses national accounting categories, whereas I used input-output data and converted these to Marxian labor values. Here, too, both are legitimate approaches, though mine is closer to Marx's original formulation. Third, I included circulating capital (that is, intermediate inputs in the input-output tables) along with fixed capital in my definition of constant capital, whereas Moseley defines constant capital as the sum of fixed-capital and business-held inventories. If the turnover period is a year, the two measures should, in principle, be identical. If not, Moseley's concept is preferable. I used intermediate inputs in my calculation of the organic composition, since they were available on a sectoral level, whereas data on business inventories held by sector were not. Fourth, Moseley includes taxes on wages as part of variable capital, whereas I included them in surplus value. The issues involved here are rather complex and devolve on the distribution of government expenditures and transfers (see Edward Wolff, 1986a) for a discussion of these issues.

Accounting Framework and the Movement of the Organic Composition. It is helpful to illustrate the importance of some of these differences by presenting comparative measures of the organic composition. The basic data used for the analysis are standard Bureau of Economic Analysis (BEA) 85 sector U.S. input-output tables and corresponding labor and capital coefficients for years 1947, 1958, 1963, 1967, 1972, and 1976.[2] All components are in constant (1958) dollars unless otherwise noted. Define:

a = Interindustry technical coefficient matrix;

X = column vector of gross domestic output (GDO) by sector;

L = row vector of employment by sector;

$N = \Sigma L$, total employment in the economy;

l = row vector of labor coefficients, where $l_i = L_i / X_i$;

k = capital coefficient matrix of the same dimensionality as a, where k_{ij} shows the amount of fixed capital of type i used per (constant) dollar of output in sector j.

Then the vector of labor values λ is given by

$$\lambda = lq, \tag{1}$$

where λ_i is interpreted as the direct plus indirect labor required per (constant) dollar of output in sector i.

Total constant capital C is defined as the labor value of the total circulating (a) and fixed (k) capital in the economy and thus given by

$$C = \lambda(k + a)X. \tag{2}$$

We can now define the economywide organic composition of capital θ as

$$\theta \equiv C/N = \lambda(k + a)X/N, \tag{3}$$

which show the ratio of total constant capital to the total labor value added per year (N).[3]

---

[1] There is a similar ambiguity in the treatment of nonprofit institutions. Many, such as private universities, research institutes, and hospitals, provide services to individuals on a fee for service basis. In this way, they behave like other capitalist enterprises. Though nominally nonprofit, they invest and expand over time. Yet, here too, there are no market forces that would cause the nonprofit sector to obtain the average rate of profit on its capital. As a result, its classification is ambiguous.

[2] See Wolff (1986b) for a complete description of the sources and methods for the input-output data for years 1947–76.

[3] I prefer this definition of the organic composition of capital to the ratio of total constant capital C to total variable capital V. The latter is more properly speaking the "value composition." It is a function of both the organic composition of capital and the rate of surplus value. To see this, define the rate of surplus value ε as $\varepsilon = (N - V)/V$. Then, the value composition ω is given by $\omega \equiv C/V = (C/N)\cdot(N/V) = \theta/(1 - \varepsilon)$.

As Moseley argues and as I have argued previously (see Wolff, 1986a), there are two forms of unproductive activity. The first consists of sectors whose output is itself unproductive. In the classification system used here, this would include the trade sectors (sector numbers 69 and 85), finance and insurance (70), real estate rentals (71), business services (73), and a portion of government services (84).[4] The second consists of unproductive activities within sectors whose output is productive. Thus, various clerical, legal, administrative, and sales workers employed within productive sectors, as well as the circulating and fixed capital used in these activities, would be counted as unproductive inputs. Though it is feasible to classify employment into productive and unproductive categories (see Wolff, 1986a for details), it is quite difficult to classify fixed and circulating capital into these two components. As a result, I will base my estimates only on a sectoral breakdown of capital and employment into productive and unproductive categories.

Between 1947 and 1976, the organic composition for all sectors increased only slightly, from 2.67 to 2.82, or by 6 percent. If only fixed capital is included in the numerator, the organic composition increased from 1.54 to 1.83, or by 19 percent over the period. The organic composition for the private economy grew by 20 percent. Finally, if we limit the organic composition to the productive sectors of the economy, then it increased from 2.83 to 3.59, or by 27 percent over the period. This last figure is still lower than Moseley's estimate that the organic composition increased by 41 percent.

---

[4] See Wolff (1986a) for a discussion of the classification of government services into productive and unproductive components.

REFERENCES

Moseley, Fred, (1986a) "The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67: A Critique and Update of Wolff's Estimates," *American Economic Review*, March 1988, *78*, 298–303.

Wolff, Edward, "The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67," *American Economic Review*, June 1979, *69*, 329–41.

Wolff, Edward N., (1986a) *Growth, Accumulation, and Unproductive Activity: An Analysis of the Postwar U.S. Economy*, New York: Cambridge University Press.

______, (1986b) "The Productivity Slowdown and the Fall in the Rate of Profit, 1947–76," *Review of Radical Political Economics*, Spring-Summer, *18*, 87–109.

NEW

Japan and the World Economy

The International Journal of Theory and Policy

Volume 1 is to be published in 1988/1989.

Editor:
Ryuzo Sato, New York University, USA

Associate Editors

Hajime Hori
Tohoku University, Japan

Takashi Negishi
University of Tokyo, Japan

Rama Ramachandran
(Coordinating Editor)
Southern Methodist University and
New York University, USA

Paul Wachtel
New York University, USA

Richard Zeckhauser
Harvard University, USA

Aims and Scope

Japan's rise in economic significance has rendered notions of Western preeminence obsolete. The Japanese economy today exerts a major impact on the world in general and the U.S. in particular. Increased economic interdependence between Japan and other trading partners raises a variety of problems and issues in economics and international business.

This Journal will publish original research in economics and international business dealing with Japan, the U.S. and other trading partners, to further understanding of the substantive problems of this increased interdependence. Accordingly, such issues as macroeconomic coordination, applied microeconomic analysis, trade friction, foreign investment, technological competition and so forth, fall under its scope. Theoretical analysis is encouraged, particularly if this leads to discussions concerning international economic policy. Empirical research is desirable, as is policy-oriented analysis that sheds light on economic and business issues affecting Japan, the U.S. and other industrialized trading partners.

It is the editor's intention that this Journal meet the growing need for comprehensive, impartial and scientific study of these issues. It is therefore particularly desirable to receive contributions from a broad-based, international community of scholars.

Call for papers

Authors of articles that fit the scope of Japan and the World Economy are invited to send their manuscripts to:

Professor Ryuzo Sato
New York University
The Center for Japan-U.S. Business
and Economic Studies
100 Trinity Place
New York, NY 10006
U.S.A.
Telephone no.: (212) 285-6050

North-Holland

In the U.S.A. and Canada:
Elsevier Science Publishers
Attn: Journal Information Center
52 Vanderbilt Avenue
New York, NY 10017, U.S.A.

In all other countries:
North-Holland
(A Division of Elsevier Science Publishers B.V.)
Attn: C. Fennes
P.O. Box 1991
1000 BZ Amsterdam, The Netherlands

NH/ECON/JNL/6191a

St. Martin's. . .Understanding Today's

Economic Issues Today

ALTERNATIVE APPROACHES

Fourth Edition

ROBERT CARSON, State University of New York, College at Oneonta

This text examines sixteen current economic issues from three different ideological perspectives—conservative, liberal, and radical. The author uses contemporary, relevant issues enhanced by charts, tables, and graphs to bring basic economic concepts and theories to life. New to this edition is a chapter on government deficits, and the chapter on environmental economics now deals with acid rain.

Paperbound. 386 pages. 1987

Also available in two separate paperbound volumes, *Macroeconomic Issues Today* and *Microeconomic Issues Today*

Instructor's Manual available.

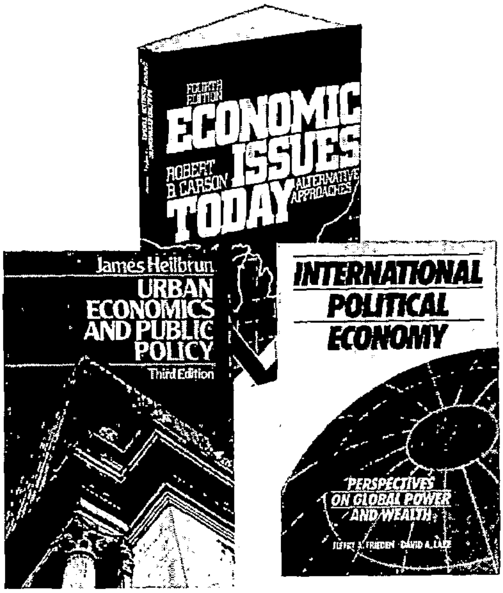
Urban Economics and Public Policy

Third Edition

JAMES HEILBRUN, Fordham University

A clear, comprehensive introduction to the study of urban economics, this text offers a balanced treatment of urban economic theory and public policy issues. The book first covers growth theory, location theory, and the system of cities; and then applies these theories to such key urban problem areas as transportation; poverty, housing, and public finance. This new edition is updated with current statistics, and the policy chapters are fully revised to reflect changes introduced during the Reagan administration.

506 pages. 1987


FOURTH EDITION
ECONOMIC ISSUES TODAY
ROBERT B. CARSON
ALTERNATIVE APPROACHES
James Heilbrun
URBAN ECONOMICS AND PUBLIC POLICY
Third Edition
INTERNATIONAL POLITICAL ECONOMY
PERSPECTIVES ON GLOBAL POWER AND WEALTH

International Political Economy

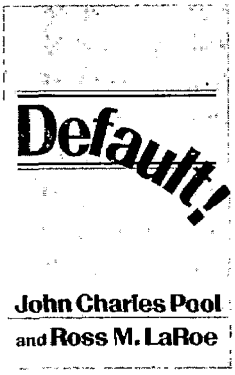
PERSPECTIVES ON GLOBAL POWER AND WEALTH

JEFFRY A. FRIEDEN and DAVID A. LAKE, both of the University of California, Los Angeles

The first comprehensive reader on the politics of international economic relations designed for use by undergraduates, this text examines the three major theoretical perspectives of international political economy—Liberalism, Marxism, and Realism—and covers a wide variety of issues, focusing on international investment, money, and trade. While the readings trace the evolution of international political economy, they also discuss current issues such as the international debt crisis.

Paperbound. 420 pages. 1987

Economy and Its Effect on Tomorrow.

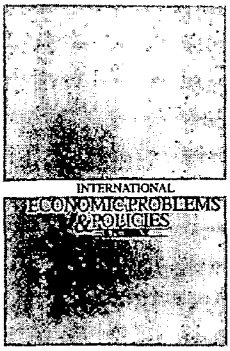


Default!

JOHN CHARLES POOL
ROSS M. LAROE, Denison University

This economics murder mystery realistically depicts the chaos that could result from the default of a Third World country on its foreign debt. Sharing in the adventures of Economics Professor Cameron Marshall as he is caught in a murderous web of international financial intrigue that includes a unilateral debt default by Mexico, students can appreciate the very real possibility of a collapse of our increasingly fragile international financial system.

Paperbound. 139 pages. Just Published



International Economic Problems and Policies

H. PETER GRAY, Rutgers University

A nontechnical, issue-oriented introduction to international economic phenomena and their influence on domestic policies, this book examines both the commercial policy (trade) and the mechanics of international payments (finance). Gray provides unique in-depth coverage of multinational corporations, focusing on such important real-world issues as the use and misuse of trade warfare and the effects of international trade on employment.

486 pages. 1987

To request an examination copy of any of these titles, please call us at our toll-free number, 800-221-7945, and ask for extension 554; or write us on your college letterhead specifying your course title, present text, and approximate enrollment. Send your request to:

St. Martin's Press

College Division•Department JR•175 Fifth Avenue•New York, NY 10010 SMP

NEW FOR 1988

Contemporary Titles in Economics from South-Western

New

ECONOMICS: A Contemporary Introduction
William McEachern, University of Connecticut, Storrs

This engaging new text provides a fresh, contemporary approach to economic principles.

Using the AS-AD framework to back *macroeconomic* theory, **ECONOMICS** covers current issues in macroeconomics such as U.S. productivity, federal budgeting, and the natural rate hypothesis.

This text's timely treatment of *microeconomics* underscores the role of time and information in production and consumption. It also reflects the increasing interest in economic institutions such as households, firms, and governments.

Numerous examples, diagrams, and case studies reinforce theory throughout the text, while the supplementary package is the most complete available.

New

MACROECONOMICS: Theory and Policy
Steven M. Sheffrin, University of California, Davis
David A. Wilton, University of Waterloo
David M. Prescott, University of Guelph

The authors of this rigorous, forthright text draw on their own experiences to bring insight and clarity to the study of policy.

The juxtaposition of the standard IS-LM curve with the PEP model provides a solid framework for analysis throughout the text.

Timely topics, such as hysteresis and the European unemployment rate, exchange rate overshooting, the new classical economics, and real business cycle theory, are covered in detail. And that coverage is rounded out by extensive discussion of budget deficits (their origins, the theories surrounding them, and Gramm-Rudman's constitutional solutions).

Newly Revised

MANAGERIAL ECONOMICS: Analysis, Problems, Cases, 3rd Edition
Lila J. Truett and Dale B. Truett
The University of Texas, San Antonio

This managerial/applied economics text offers solid coverage of central issues—such as profit maximization—and devotes the first five chapters to an explanation of all the tools and decision criteria (revenue, demand, production, cost, profit-maximization) that your students need to analyze problems later in the text. The expanded supplementary package now includes *The Decision Assistant*, a software applications package which provides students with "hands-on" analysis in break-even analysis, linear programming, price discrimination and more.

Newly Revised

MACROECONOMICS: Intermediate Theory and Policy, 2d Edition
William J. Boyes, Arizona State University

This student-oriented text is written in an accessible style that clarifies concepts and puts ideas in perspective.

It includes a thorough discussion of theory, as well as basic macro policy issues and background information—which makes it suitable for M.B.A. as well as undergraduate programs.

Material has been revised and updated according to feedback from reviewers—and a new chapter providing a comparison of the basic elements of microeconomics and macroeconomics has been added.

A Study Guide and complete Instructor's Manual provide valuable aid to you and your students.

For more information on these and other economics texts, contact the South-Western representative in your area.

SOUTH-WESTERN
COLLEGE DIVISION
5101 Madison Road • Cincinnati, Ohio 45227

ySTAT™

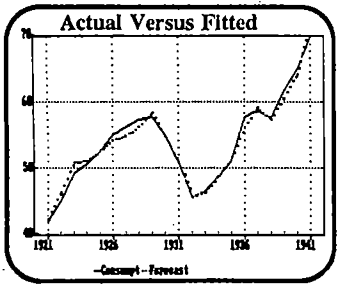
For IBM Personal Computers

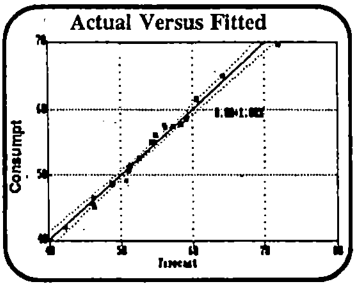
Integrates
SPREADSHEET + STATISTICS + FORECASTING + GRAPHICS
and is Easy to Use.

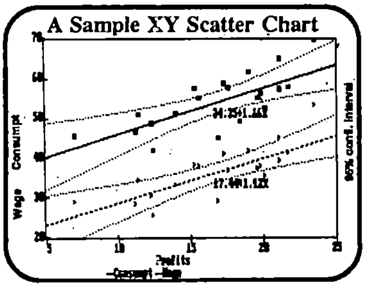
- ySTAT is the only full-fledged spreadsheet statistical program - a fast, powerful, easy-to-use stand-alone program.
- **Menu-driven** format as in Lotus 1-2-3 - offers many functions that are not available even in 1-2-3, such as vector formulas, lagged variables, dummy variables, moving averages, and interpolation of missing values.
- You can read, enter, edit, copy, move, and sort the data as well as group and partition the data for analysis right on the spreadsheet. A total of 40 function keys for one-keystroke operations.
- **Ten statistical function keys:** to tally sums and means; subtract variable means or seasonal means; standardize the data; aggregate seasonal data; generate seasonal indicators; perform moving averages, linear trend forecasts, and so on.
- **ySTAT reads 1-2-3 and Symphony worksheet files directly,** and any free-format or fixed-format textfile generated by other PC programs or downloaded from mainframe. The data can be numeric, alphabetic, including missing values.
- Mean, median, quartiles, skewness, mode, t-test, frequency distribution, correlation, crosstabs, analysis of variance.
- Multiple **regression** (OLS), standardized regression, weighted least squares, two-stage least squares, polynomial regression, and Cochrane-Orcutt method. Including diagnostic residual analysis output (see sample at right).
- **Pooling** of cross-section and time-series data.
- Two-sample difference of means test, **nonparametric tests** (Kolmogorov-Smirnov test, Wald-Wolfowitz runs test, Mann-Whitney test, Wilcoxon matched-pairs signed-ranks test) and Spearmans's and Kendall's rank-order correlations.
- **Time-series models:** autoregressive model, moving average model, and Box-Jenkins model.
- **Nonseasonal forecasts:** linear trend, quadratic trend, polynomial model, sinusoidal model; simple, double and triple exponential smoothings.
- **Seasonal forecasts:** exponential smoothings with seasonal indicators or with trigonometric functions; Winters' additive and multiplicative procedures.
- **LOGIT, PROBIT, TOBIT,** and Weighted Probit regressions: these models are twice as fast as Gauss program and faster than other programs and are much easier to use.

High-resolution **color charts** : the actual and forecast values are graphically displayed as a line chart and also as an XY chart, including a regression line and its 95% confidence interval.

```
                         OLS  -- DEPENDENT VARIABLE: Consumpt

   RIGHT-HAND           ESTIMATED            STANDARD        T_STATISTIC        PROB.
    VARIABLE            COEFFICIENT           ERROR

 1  Profits            0.192934381     (     0.09121)    T=     2.11527        0.049
 2      P-1            0.089884898     (     0.09065)    T=     0.99150        0.335
 3     W+W'            0.796218750     (     0.03994)    T=    19.93342        0.000
 4 Constant           16.236800272     (     1.30270)    T=    12.46382        0.000

  SAMPLE SIZE(  1 to   21) =            21          (DF=17)
  SUM OF SQUARED RESIDUALS =            17.879449
           VARIANCE (MSE) =             1.051732
 STANDARD ERROR (ROOT MSE) =            1.025540
                R-SQUARED =             0.981008
       ADJUSTED R-SQUARED =             0.977657
   F-STATISTIC(  3,   17) =           292.707595 (p=0.0004)
         SUM OF RESIDUALS =            -0.000000
   DURBIN-WATSON STATISTIC =            1.367474

Analysis of Variance|            Source |     SUM SQ    | DF  |  MEAN SQ   |
                    |-------------------|---------------|-----|------------|
                    |Due to Regression  |      323.550| 3|    307.850|
                    |          Residual |       17.879| 17|      1.052|
                    |             Total |      341.430| 20|     47.071|
                    |-------------------|---------------|-----|------------|

RESIDUAL ANALYSIS -Mean: -0.000   Adj. RMSE: 0.346   Mean Abs. % Err: 1.283

                          AUTOCORRELATIONS

LAG   COEF    T-VAL   -+----+----+----+----+----0----+----+----+----+----+-  LAG
  1  0.181     0.83   |                         |>>>>>>>>>>                 |    1
  2 -0.059    -0.27   |                      <<<|                           |    2
  3 -0.033    -0.15   |                       <<|                           |    3
  4 -0.157    -0.69   |                <<<<<<<<<|                           |    4
  5 -0.411    -1.81   |<<<<<<<<<<<<<<<<<<<<<<<<<|                           |    5
                      -+----+----+----+----+----0----+----+----+----+----+-  LAG
    Ljung-Box statistic (chi-square 4 DF): 6.698 (p=0.1527)

                       PARTIAL AUTOCORRELATION

LAG   COEF    T-VAL   -+----+----+----+----+----0----+----+----+----+----+-  LAG
  1  0.181     0.83   |                         |>>>>>>>>>>                 |    1
  2 -0.095    -0.44   |                   <<<<<<|                           |    2
  3 -0.004    -0.02   |                         |                           |    3
  4 -0.162    -0.74   |               <<<<<<<<<<|                           |    4
  5 -0.378    -1.73   |<<<<<<<<<<<<<<<<<<<<<<<<<|                           |    5
                      -+----+----+----+----+----0----+----+----+----+----+-  LAG

                  ACTUAL versus FITTED VALUES and RESIDUALS

SEQ    Actual   Fitted min=41.900                     71.873=max  Error
        (*)      (+)   -+---+---+---+---+---+---+---+---+---+-           +---+---+
  1    41.90    42.22 |@                                     |  -0.324 |  <:    |
  2    45.00    46.25 |  *+                                  |  -1.250 |<<<:    |
  3    49.20    50.77 |      * +                             |  -1.566 |<<<:    |
  4    50.60    51.09 |        @                             |  -0.494 |  <:    |
  5    52.60    52.59 |          @                           |   0.008 |   :>   |
  6    55.10    54.23 |             +*                       |   0.869 |   :>>  |
  7    56.20    54.86 |             +  *                     |   1.338 |   :>>> |
  8    57.30    56.25 |               +*                     |   1.055 |   :>>  |
  9    57.80    58.39 |                *+                    |  -0.589 | <<:    |
 10    55.00    54.72 |             @                        |   0.282 |   :>   |
 11    50.90    51.13 |        @                             |  -0.230 |  <:    |
 12    45.60    45.92 |  *+                                  |  -0.322 |  <:    |
 13    46.50    46.18 |   @                                  |   0.322 |   :>   |
 14    48.70    48.76 |      @                               |  -0.058 |  <:    |
 15    51.30    51.33 |         @                            |  -0.035 |  <:    |
 16    57.70    56.08 |               +*                     |   1.616 |   :>>> |
 17    58.70    59.14 |                  *+                  |  -0.436 |  <:    |
 18    57.50    57.29 |                 @                    |   0.210 |   :>   |
 19    61.60    60.61 |                    +  *              |   0.989 |   :>>  |
 20    65.00    64.21 |                        +*            |   0.785 |   :>>  |
 21    69.70    71.87 |                               *    +|  -2.173 |<<<<:   |
        (*)      (+)   -+---+---+---+---+---+---+---+---+---+-           +---+---+

          LIST OF OUTLIERS - STUDENTIZED RESIDUALS GREATER THAN 1.50

    SEQ     Actual    Fitted     Residual  Std. Err  Studentized Probability
                                                      Residuals     DF=17
    ----    ------    -------    --------  --------  ----------  ----------
      3     49.20     50.77      -1.566     0.967     -1.619       0.124
     16     57.70     56.08       1.616     0.958      1.687       0.110
     21     69.70     71.87      -2.173     0.780     -2.787       0.013
    ----    ------    -------    --------  --------  ----------  ----------

(Note: this sample output shows some of the options for regression diagnostics.)
```







MING TELECOMPUTING INC.

23 Oak Meadow Road, P.O. Box 101, Lincoln Center, MA 01773, U.S.A. (617) 259-0391

Order: ( ) ySTAT for $395. ( )ySTAT/Medical for $395.
( ) Trial Disk and information/sample output for $5.
( ) Information/sample output.
System: ( ) IBM PC. ( )XT. ( )AT. ( )System 2 Model____
( ) Other system ____________________
Options: ( ) with 8087 or 80287 or 80387 math co-processor.
( ) IBM CGA ( ) EGA ( ) Hercules Mono. Card
Type of printer or plotter ____________________

Payment: ( ) Check or money order.
( ) U.S. university or governmental purchase order.
( ) Visa. ( ) MC. Card No. ____________________
Name ____________________ Expir. date ___/___
Address ____________________
____________________ Telephone: ( ) __________

IBM is a registered trademark of International business Machines Corp.; 1-2-3 of Lotus Development Corp.; Gauss of Aptech Systems, Inc.

Beginning a second decade of publishing service to the academic community...

POLITICAL AND SOCIAL ECONOMY

A series edited by C. ADDISON HICKMAN
and MICHAEL P. SHIELDS

Titles published in the series:

CONGLOMERATES AND THE EVOLUTION OF CAPITALISM
By Charles R. Spruill

AN ECONOMIC ANALYSIS OF DEMOCRACY
By Randall G. Holcombe

ECONOMIC THOUGHT AND SOCIAL CHANGE
By J. Ron Stanfield

ORIGINS OF ECONOMIC THOUGHT AND JUSTICE
By Joseph J. Spengler

THE POLITICAL ECONOMY OF THE URBAN GHETTO
By Daniel R. Fusfeld and Timothy Bates

PUBLIC FINANCE AND THE POLITICAL PROCESS
By Randall G. Holcombe

TOWARD SOCIAL ECONOMY
By Howard R. Bowen

Southern Illinois University Press solicits worthy manuscripts in economics, political science, sociology, and history. Please address proposals to: Editorial Director, SIU Press, P.O. Box 3697, Carbondale, IL 62901.

SOUTHERN ILLINOIS UNIVERSITY PRESS
Dept. AE38, P.O. Box 3697, Carbondale, IL 62902-3697

New York University Press

Political Economy and International Money

Essays by John Williamson

CHRIS MILNER, Editor

320 pages/ISBN: 0-8147-5434-1/$45.00

Arrow and the Foundations of the Theory of Economic Policy

Arrow and the Ascent of Modern Economic Theory

GEORGE R. FEIWEL, Editor

Foundations: *832 pages/ISBN: 0-8147-2583-X/$75.00*

Ascent: *768 pages/ISBN: 0-8147-2582-1/$75.00*

IRM Directory of Statistics of International Investment and Production

JOHN DUNNING and JOHN CANTWELL

820 pages/ISBN: 0-8147-1783-7/$125.00

Public Finance in a Democratic Society

Volume I: Public Finance in a Democratic Society

Volume II: Fiscal Doctrine, Growth and Institutions

RICHARD A. MUSGRAVE

Volume I: *392 pages/ISBN: 0-8147-5428-7/$65.00*

Volume II: *400 pages/ISBN: 0-8147-5429-5/$65.00*

The Set: *ISBN: PUBVOLSET/$130.00*

Human Desire and Economic Satisfaction

Essays on the Frontiers of Economics

TIBOR SCITOVSKY

210 pages/ISBN: 0-8147-7862-3/$45.00

NEW YORK UNIVERSITY PRESS

Washington Square, New York, NY 10003

Is your publisher using his best lines to get yours?

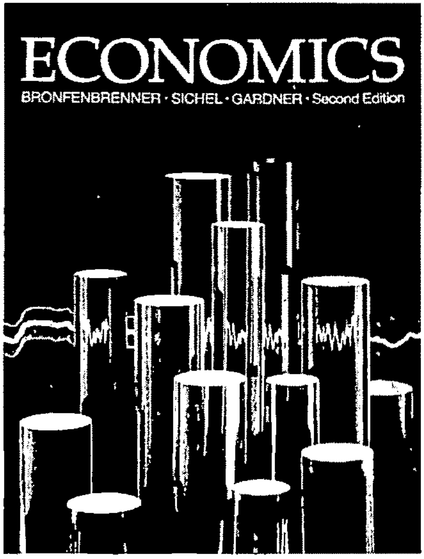
Just remember, he's rehearsed them on hundreds of authors before you.

And there is a more professional way to publish your work. Talk to us at South-Western. You'll get honest, personal attention. And the respect you deserve.

From no-nonsense, in-house editors. Outside reviewers you can trust. And an aggressive sales force that will get to know your book before they sell it.

Call (513) 527-6384. We'll give you the same kind of time and attention you gave your manuscript.

SOUTH-WESTERN
COLLEGE DIVISION



THE RATIONAL CHOICE IS THE NATIONAL CHOICE.

Economics, Second Edition
Martin Bronfenbrenner
Aoyama Gakuin University, Japan
Emeritus, Duke University
Werner Sichel and **Wayland Gardner**
Both of Western Michigan University
Complete Hardcover Edition • 828 pages
Two-volume Paperback Edition:
Macroeconomics • 460 pages
Microeconomics • 556 pages
Study Guide and Test Bank by **Rose Pfefferbaum,** Mesa Community College
Instructor's Manual • Transparency Masters
Color Transparencies • MicroStudy: Computerized Study Guide • MicroTest: Computerized Test Bank • Call-in Test Service • Color Graphics Software
GPA: Grade Performance Analyzer
PC Economics: Interactive Graphing Software and Workbook • 1987

For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
University Sq., 101 Campus Dr.,
Princeton, NJ 08540

Microeconomics.
It's not a spectator sport.

Most economics texts start with buzz words...telling students how economists *talk* instead of how they *think.* It's a bad start, because it puts students *outside* the process of economic analysis.

Now there's a textbook that plainly describes how economists think and how they analyze problems. It lets students in on the economic method right at the start, and challenges them, again and again, to *use* it.

Announcing Eaton and Eaton: MICROECONOMICS

Because to learn economics is to *do* economics.

Find out how good a textbook can be. Write for your free examination copy.

W. H. FREEMAN AND COMPANY
The book publishing arm of *Scientific American*
41 Madison Avenue, New York, NY 10010

Oxford

Again available in paperback—by the winner of the 1987 Nobel Prize in Economics

Growth Theory
An Exposition

ROBERT M. SOLOW, *Massachusetts Institute of Technology*

With American productivity growing at ever-smaller rates and international stock markets in spectacular turmoil, this distinguished book by one of the greatest living economists is more timely than ever. Organized around the famous Harrod-Domar conditions for equilibrium in a growing economy, *Growth Theory: An Exposition* presents a concise survey of the modern macroeconomic theory of growth, showing how the character of full-employment growth equilibrium changes and enlarges as more complicated assumptions are made about technology, invention, saving, and the existence of monetary assets. The book explores general policy issues and includes an elementary exposition of the problem of optimal paths of capital accumulation, and the discounting of the benefits of public investments.

1970 (paper 1988) 128 pp.
paper $10.95

Prices are subject to change.
To request an examination copy, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to: College Humanities & Social Sciences Marketing Dept.

Oxford University Press
200 Madison Avenue • New York, NY 10016

North-Holland
New Books in Economics

Macroeconomic Medium-Term Models in the Nordic Countries

Edited by **O. Bjerkholt** and **J. Rosted**

Contributions to Economic Analysis, 164

1987 x+324 pages
Price: US $79.00/Dfl. 150.00
ISBN 0-444-70262-8

This is the first major joint presentation of macroeconomic models of Denmark, Finland, Norway and Sweden. Its origin is a selection of the presentations at a seminar on macroeconomic modelling in the Nordic countries held in Lybgby, Denmark, 8–9 October 1984, sponsored by the Nordic Ministries of Finance.

Economic-Ecological Modeling

Edited by **L.C. Braat** and **W.F.J. van Lierop**

Studies in Regional Science & Urban Economics, 16

1987 xii+330 pages
Price: US $79.00/Dfl. 150.00
ISBN 0-444-70298-9

This book is about mathematical models for environmental and resource policy and management. The authors present an overview of the theory, methods techniques and experience relevant to the analysis of problems on the interface of society and its natural environment.

Capital Income Taxation and Resource Allocation

By **H.-W. Sinn**

Studies in Mathematical and Managerial Economics, 35

1987 xiv+414 pages
Price: US $55.00/Dfl. 175.00
ISBN 0-444-70208-3

This monograph investigates the intersectoral, international, and intertemporal allocation effects of alternative systems of capital income taxation characterized by different degrees of integration between corporate and personal taxation, depreciation rules, provisions for interest deductibility, and the like. The systems studied include those of the OECD countries as well as proposed systems advocated by various authors and tax committees.

Regional Labour Markets

Analytical Contributions and Cross-National Comparisons

By **M.M. Fischer** and **P. Nijkamp**

Contributions to Economic Analysis, 168

1987 xviii+500 pages
Price: US $100.00/Dfl. 190.00
ISBN 0-444-70323-3

The current state of spatial labour market research is presented in this book, by addressing analytical issues, by identifying and evaluating current trends on urban and regional labour markets and by comparing the operations of spatial labour markets in various industrialised countries. The central focus of the volume is on a cross-national comparison of regional labour markets in 15 countries, based on a joint research effort of teams from these countries.

Japanese Industrial Performance

By **K. Uno**

1987 xx+440 pages+a diskette
Price: US $94.75/Dfl. 180.00
ISBN 0-444-70274-1

This analysis of the performance of Japanese industry covers the period 1955 to the 1980s. The main purposes of the work are to provide a systematic analysis of the economy disaggregated to 25 industrial sectors and to explore the feasibility of large scale industry modelling, from the viewpoint of both the availability and reliability of statistical data. The book covers current policy issues as well as historical development.

North-Holland

In the U.S.A. and Canada:
Elsevier Science Publishing Co. Inc.,
P.O. Box 1663, Grand Central Station,
New York, NY 10163, U.S.A.

In all other countries:
Elsevier Science Publishers,
Book Order Department,
P.O. Box 211,
1000 AE Amsterdam, The Netherlands.

US $ prices are valid only in the USA and Canada. In all other countries the Dutch Guilder (Dfl.) price is definitive. Customers in the Netherlands, please add 6% B.T.W. In New York State applicable sales tax should be added. All prices are subject to change without prior notice.

NH/ECON/BK/0970

THE ECONOMICS INSTITUTE

Gateway to Successful
Master's and Doctoral Degree Studies
in Economics, Agricultural Economics, Business and Administration

Since its establishment in 1958 under the sponsorship of the American Economic Association, the Institute has provided specialized preparatory training and orientation to over 6,000 students from over 120 countries en route to over 320 universities in the United States and other English-speaking countries.

PROGRAM FEATURES

- Six to forty-six week individualized programs.
- Intensive and comprehensive review or supplementary preparatory training in:
  English, Computer Usage, Economic Theory, Mathematics, Statistics, Accounting, Management, Finance and Marketing.
- Standardized test preparation (TOEFL, GRE, GMAT).
- Orientation to U.S. campus and community life.
- University placement assistance services.
- Certificate and Diploma Programs for short-term professional trainees.

END RESULTS for

- **Associated universities:** Improved admissions procedures and a source of additional high-quality foreign students.
- **Foreign students:** Greater accessibility to U.S. universities.
- **Sponsoring organizations:** Reduced total costs.
- **And for all three:** Better grade performance records in degree programs and more rapid completion of degree requirements.

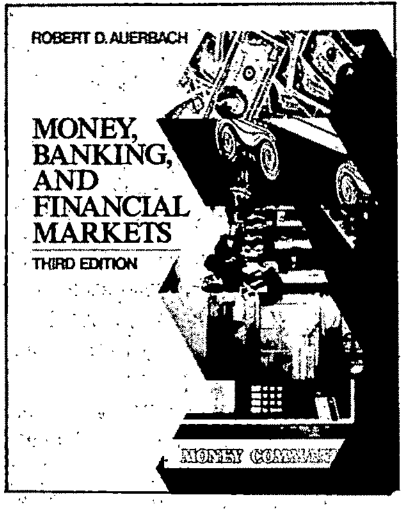
ORGANIZATION

AEA Policy and Advisory Board
Edwin S. Mills, Princeton University, Chairman
Lance E. Davis, California Institute of Technology
W. Lee Hansen, University of Wisconsin, Madison
Joseph Havlicek, Ohio State University
Teh-wei Hu, University of California, Berkeley
Samuel A. Morley, Vanderbilt University
Dwight H. Perkins, Harvard University
Stefan H. Robock, Columbia University

Director
Wyn F. Owen, University of Colorado

The Economics Institute, 1030 13th Street, Boulder, Colorado, 80302, USA;
Telephone: (303) 492-3000; Telex: 450385 ECONINST BDR

MACMILLAN
PREPARES YOUR STUDENTS FOR
FUTURE ECONOMIC DEMANDS



NEW IN 1988

MONEY, BANKING, AND FINANCIAL MARKETS
Third Edition
ROBERT D. AUERBACH,
University of California–Riverside
816 pages Instructor's Resource Manual, Study Guide, MICROTEST, a computerized test generator

INTRODUCTION TO ECONOMETRICS
G. S. MADDALA, University of Florida
464 pages Solutions Manual

ECONOMIC STATISTICS AND ECONOMETRICS
Second Edition
THAD W. MIRER, SUNY Albany
416 pages Instructor's Manual

INDUSTRIAL ECONOMICS
STEPHEN MARTIN, Michigan State University
576 pages Instructor's Manual

ADOPTION DEADLINE NEAR? For urgent adoption consideration only, dial toll-free (800) 428-3750 during our business hours 8:30 AM–4:30 PM EST (In Indiana, call (317) 298-5513). *For less immediate needs, please write:*

MACMILLAN PUBLISHING COMPANY
COLLEGE DIVISION • 866 THIRD AVENUE • NEW YORK, NY 10022

For your principles course...
New from Scott, Foresman
Little, Brown for 1988

Economics
Second Edition

Robert B. Ekelund, Jr.,
Auburn University
Robert D. Tollison,
George Mason University

This mainstream, introductory text provides clear, balanced, and comprehensive coverage of modern economics. By drawing upon an array of interesting applications and important policy issues, the text puts economic theory in terms your students will understand and appreciate.

New to the Second Edition:

- New Chapter 33, "Fiscal Policy, Taxation, and the Debt," has been added to the macro section
- Chapter 19, "The Microeconomics of Taxation," has been greatly expanded to include new material on the 1986 tax reform
- The macro section has a more fully integrated aggregate supply/aggregate demand model and offers earlier treatment of monetary theory
- **Available Now,** in single-volume and split editions, with complete supplement package

Introductory Economics

Marilu McCarty,
Georgia Institute of Technology

- Economics Tool Kit, an integrative computer software package, comes with each copy of this text at no extra cost; various text passages and problem sections are keyed to the software
- Unique macro chapter on the role of financial intermediation, including financial futures
- Strong public policy flavor (e.g., in health economics) and applications-oriented presentation in non-technical manner
- **Available Now,** in single-volume and split editions, with complete supplement program

Principles of Economics

Third Edition

Roy J. Ruffin / Paul R. Gregory
University of Houston

- New open economy theme throughout the macro section
- New Keynesian work on expectations, as well as Neoclassical updates
- Integration of multiyear contracting, layoff unemployment, and contract markets as sources of inflexibility
- 150 new boxed examples (e.g., Citibank write-off of Third World bad loans, search methods, role of overlapping contracts, banking concentration in Canada, taxing real estate interest rates, etc.)
- Still the only basic text with a chapter on information economics—updated with new applications
- Acclaimed by students as easy to read and to study; praised by instructors as the superb blend of new research and traditional topic coverage
- **Available Now,** in single volume and split editions, with complete supplement package, including new software program for the IBM PC

Economics Principles of Political Economy

Third Edition

Daniel R. Fusfeld,
The University of Michigan

- An extensive update of the only text to offer a systematic introduction to economics with a historical, institutionalist orientation
- New early chapter on the international economy
- Updated coverage of recent antitrust policies; problems with the "Deficit Twins"; latest studies of income/wealth distribution; the tax reform of 1986
- **Available Now.** *864 pages, hardbound, with complete supplement program*

For further information write Meredith Hellestrae, Department SA—AER
1900 East Lake Avenue Glenview, Illinois 60025

Scott, Foresman / Little, Brown College Division

New and Noteworthy

FISCAL POLICIES AND THE WORLD ECONOMY

An Intertemporal Approach

Jacob A. Frenkel and Assaf Razin

Frenkel and Razin analyze the consequences of both the budget deficit and the trade deficit on the world economy within a sound theoretical framework that uses the modern intertemporal approach and is unique in its ability to analyze the domestic and international repercussions of changes in fiscal policies.

$24.00

Recipient of the National Science Foundation Waterman Prize

UNDERSTANDING UNEMPLOYMENT

Lawrence H. Summers

Lawrence Summers explores new theories of joblessness that could eventually explain why unemployment remains high despite relatively healthy economic growth. He provides the sound empirical base that is essential to the design of effective policies for combating this pressing social problem.

$22.50

DOMESTIC AND INTERNATIONAL BANKING

Mervyn K. Lewis and Kevin T. Davis

Lewis and Davis investigate the theory and practice of domestic and international banking and finance. They provide general background on payments systems, Eurocurrency markets, bank safety, and depositor protection, and also trace parallels between operations of banks and other financial institutions, particularly insurance companies.

$25.00

PRIVATE ANTITRUST LITIGATION

New Evidence, New Learning

edited by Lawrence J. White

Is private antitrust litigation out of control, encouraging frivolous suits and deterring companies from pursuing innovative manufacturing, organization, and distributional techniques? Or is it a fair and useful system, particularly during periods when government antitrust enforcement is lax and pro-business? Using a unique collection of data on more than 2,350 antitrust cases filed in five districts between 1973 and 1983, prominent scholars analyze the key issues involved in reform proposals.

$40.00

The MIT Press

55 Hayward Street, Cambridge, MA 02142

Now in Paperback

THE ECONOMICS OF REGULATION

Principles and Institutions

Alfred E. Kahn

The first comprehensive integration of the economic theory and institutional practice of economic regulation. With a lengthy new introduction, this edition of *The Economics of Regulation* clearly remains the standard work in the field, a starting point and reference tool for anyone working in regulation.

$25.00 paper ($45.00 cloth)

LIVES OF THE LAUREATES

Seven Nobel Economists

edited by William Breit and Roger W. Spencer

A condensed and personalized history of modern economic thought, with some of the most eloquent and important contributors to that history as guides.

$6.95 paper ($13.95 cloth)

THE INVISIBLE LINK

Japan's Sogo Shosha and the Organization of Trade

M. Y. Yoshino and Thomas B. Lifson

"For foreigners wanting to know what makes the point men in Japan's external trade tick, this is a valuable book." — *Business Week*

$9.95 paper ($22.50 cloth)

JAPAN'S RESHAPING OF AMERICAN LABOR LAW

William B. Gould IV

"The first major contribution on comparative aspects of Japanese and American labor practices outside of the legal and trade literature, the book goes a long way toward sorting out some of the answers to the question, 'How do they do it?' as well as that question's inevitable follow-up: 'What can we (or should we) learn from them?'" — Jennifer Friesen, *Michigan Law Review*

$9.95 paper ($24.95 cloth)

1986 American Association of Publishers Outstanding Book in Business, Management, and Economics

SUPERFAIRNESS

Applications and Theory

William J. Baumol

With his characteristic acuteness and lucidity, William Baumol tackles the problem of equity considerations in welfare economics by applying the novel "superfairness" criterion to the distribution of resources, product, income, and wealth that arises from economic decisions.

$12.50 paper ($22.00 cloth)

Money, Interest, and Banking in Economic Development

Maxwell J. Fry

Combining theory, empirical evidence, institutional description, and policy analysis, Maxwell Fry reviews and evaluates the rôle of monetary economics in developing economies.

"Fry's book will make a big impact on our profession—not only for academic use but as a major contribution to research in the area of development finance."
—Ronald I. McKinnon, Stanford University

The Johns Hopkins Studies in Development
Vernon W. Ruttan and T. Paul Schultz, Consulting Editors

$37.50 *hardcover* **$14.95** *paperback*

Working but Poor
America's Contradiction

Sar A. Levitan and Isaac Shapiro

The American ethos equates work with material progress, but nine million Americans labor in poverty. *Working but Poor*—the first book in more than a decade to focus on this frequently overlooked segment of the poor population—profiles the working poor, evaluates government efforts in their aid, and calls for a series of reforms to reverse the recent sharp increase in their number.

$22.50 *hardcover* **$7.95** *paperback*

Published for the World Bank—

Cost Analysis in Education
A Tool for Policy and Planning

Philip H. Coombs and Jacques Hallak

This book presents a practical approach to measuring and estimating educational costs and to using that information to make the best possible use of scarce resources. Although not a panacea for the financial crisis currently facing education systems throughout the world, creative cost analysis is a powerful and indispensable tool for educational policy makers, planners, and managers.

EDI Series in Economic Development

$14.50 *hardcover*

THE JOHNS HOPKINS UNIVERSITY PRESS
701 West 40th Street, Suite 275, Baltimore, Maryland 21211

American Economic Association

Summer Minority Program

at Temple University

June 1, 1988–July 27, 1988

A New Economics Program with Exciting Features

An Attractive Stipend–plus Books,
Room, Board, Tuition and Travel are
Available to Qualified Students.

For More Information Contact:

DEPARTMENT OF ECONOMICS
RITTER HALL ANNEX
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122

(215-787-8887)

AMERICAN ECONOMIC ASSOCIATION

1988 ANNUAL MEMBERSHIP RATES

Membership includes:

—a subscription to *The American Economic Review* (quarterly) plus *Papers and Proceedings*, the *Journal of Economic Literature* (quarterly) and the *Journal of Economic Perspectives* (quarterly).

● Regular members with annual incomes of $30,000 or less	$38.50
● Regular members with annual incomes above $30,000 but no more than $40,000	$46.20
● Regular members with annual incomes above $40,000	$53.90
● Junior members (available to registered students for three years only). Student status must be certified by your major professor or school registrar	$19.25
● In Countries other than the U.S.A., Add $16.00 to cover postage.	
● Family members (persons living at the same address as a regular member, additional memberships without subscription to the publications of the Association)	$7.70

Please enter my subscription for the following period:

☐ **Jan.-Dec.** ☐ **April-March** ☐ **July-June** ☐ **Oct.-Sept.**

First Name and Initial **Last Name** **Suffix**

Address Line 1

Address Line 2

City

State or Country **Zip/Postal Code**

MAJOR FIELDS (TWO ONLY)

LIST FIELDS WITH WHICH YOU CURRENTLY IDENTIFY. SELECT FIELD CODE FROM *JEL*, "Classification System for Books."

☐ ☐

Please type or print information above. Please pay with a check or money order payable in United States Dollars. Canadian and foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Please note: It is the policy of the Association, not to refund membership payments.

Endorsed by (AEA member) ______

Below for Junior Members Only

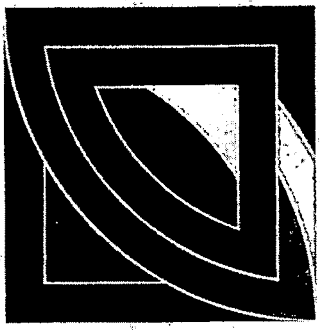
I certify that the person named above is enrolled as a student at ______

Authorized Signature

PLEASE SEND WITH PAYMENT TO:

AMERICAN ECONOMIC ASSOCIATION
1313 21ST AVENUE SOUTH, SUITE 809
NASHVILLE, TENNESSEE 37212-2786
U.S.A.

McKenzie
ECONOMICS



BENEFIT
FROM
THE EXPERIENCE.

Economics
Richard McKenzie, Clemson University
Complete Hardcover Edition: 847 pages
Macroeconomics: paperback • 477 pages
Microeconomics: paperback • 569 pages
Study Guide • MicroStudy: Computerized Study Guide • Instructor's Manual
Test Bank • MicroTest: Computerized Test Bank • Call-in Test Service
GPA: Grade Performance Analyzer
Transparencies • Computer Graphics Package • 1986

Combining experience in national policy-making and in teaching, Richard McKenzie offers a textbook that presents economic principles in terms everyone can understand.

Boxed "Perspectives on Economics"—many written by specialists especially for this text—focus on today's critical issues.

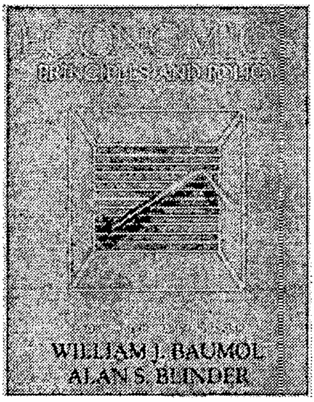
For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
University Sq., 101 Campus Dr.,
Princeton, NJ 08540

HBJ
COMMUNIQUE

ABSOLUTE ADVANTAGE for your students.



Many Principles of Economics textbooks clamor for your attention and favor at this time of year. Yet only one book clearly stands out from “the pack” this spring. That book is

ECONOMICS
Principles and Policy
Fourth Edition
WILLIAM J. BAUMOL and ALAN S. BLINDER
Hardcover. 928 pages. Also available in two-volume paperback format. Complete ancillary package available.

ECONOMICS: Principles and Policy has been setting new standards for instructional excellence through three editions with coverage focused on critical economic ideas of lasting significance, an approach that transcends the encyclopedia style of books that preceded it. From the *First Edition,* the text has employed aggregate supply/aggregate demand analysis to provide students with a much clearer explanation of today's complex economic situation. Now the just-published *Fourth Edition* expands upon these many outstanding features in a new, thoroughly updated and carefully revised presentation that places expanded emphasis on the world economy and America's place in it. This new world economic emphasis is reflected in extensive revision and restructuring of topics, not simply cosmetic “window dressing.”

Beyond the new emphasis on the world economy, a variety of other key changes have been made in the Fourth Edition:

- revised discussion of elasticity to reduce technicalities
- clarified discussion distinguishing scale economics from declining average cost
- fuller discussion of the differences among short and long runs and fixed and sunk costs
- new discussion of corporate takeovers

You probably have an examination copy of the Fourth Edition of Baumol & Blinder already. If you do not, call our Chicago Sales Office (312-647-8822) or your HBJ Representative and a copy will be *RUSHED* to you.

HBJ

HARCOURT BRACE JOVANOVICH, Inc.
College Sales Office
7555 Caldwell Avenue
Chicago, IL 60648

AEA sponsored Group Life Insurance for you and your family—at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past nine years, insured members received credits on their April 1 semiannual payment notices averaging 40% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program
1255 23rd Street, N.W.
Washington, D.C. 20037

J-1

Please send me more information about the AEA Life Insurance Plan.

Name ____________________ Age______

Address ____________________

City________________ State______ Zip ______

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

New in April

The American Enterprise Institute Trade in Services Series

Claude E. Barfield, Series Coordinator

When Countries Talk
International Trade in Telecommunications Services
Jonathan David Aronson and **Peter F. Cowhey**
0-88730-284-X 368 pages $29.95

International Trade in Services
Overview and Blueprint for Negotiations
Geza Feketekuty
0-88730-241-6 240 pages $26.95

Deregulation and Globalization
Liberalizing International Trade in Air Services
Daniel M. Kasper
0-88730-233-5 120 pages $26.95

International Trade in Construction, Design, and Engineering Services
James R. Lee and **David Walters**
0-88730-235-1 160 pages $29.95

International Trade in Business Services
Accounting, Advertising, Law, and Management Consulting
Thierry J. Noyelle and **Anna B. Dutka**
0-88730-237-8 184 pages $26.95

Global Competition in Financial Services
Market Structure, Protection, and Trade Liberalization
Ingo Walter
0-88730-234-3 328 pages $29.95

International Trade in Ocean Shipping Services
The United States and the World
Lawrence J. White
0-88730-238-6 144 pages $26.95

International Trade in Films and Television Programs
Steven S. Wildman and **Stephen E. Siwek**
0-88730-240-8 224 pages $29.95

Also of Interest

International Collaborative Ventures in U.S. Manufacturing
David C. Mowery, editor
0-88730-221-1 464 pages $34.95

Special limited-time offer. Order all eight books in the series before April 30, 1988 and receive a 20% discount.

YES! Please send me:
____ **Telecommunications Services/6612576 $29.95**
____ **Overview/6612832 $26.95**
____ **Air Services/6612584 $26.95**
____ **Construction, Design, and Engineering/6612600 $29.95**
____ **Business Services/6612352 $26.95**
____ **Financial Services/6612394 $29.95**
____ **Ocean Shipping Services/6612428 $26.95**
____ **Films and Television Programs/6612436 $29.95**
____ **International Collaborative Ventures/6612022 $34.95**

☐ Payment enclosed ☐ Bill me
Charge my ☐ MC ☐ VISA ☐ AMX

Card no. ________________ Exp. date ________
Signature ________________________
Send to: ________________________
____
________________ Zip ________

BALLINGER
PUBLISHING COMPANY
Harper & Row
Order Department
2350 Virginia Avenue, Hagerstown, MD 21740
(800) 638-3030

My state sales tax $________
Subtract 20% if ordering series ________
Postage and handling $2.75/order $________
***Prepaid orders are postage free!**
TOTAL $________

Prices subject to change. All orders subject to credit approval. U.S. funds only. If you order by phone, tell the operator your order code is **AAER388**

Timely Economic Analyses

from Lexington Books

NEW TEXT

The U.S. Economy Demystified

Revised Edition

Albert T. Sommers, The Conference Board with Lucie R. Blau

A lucid explanation of our complex economic system. It builds a framework within which to insert new information and sense the probable range of future developments.

1988 0-669-17385-1 192 pages $13.95 (paper)
0-669-17383-5 $19.95 (cloth)

NEW TEXT

Public Sector Labor Relations

Analysis and Readings

David Lewin, Columbia University, Peter Feuille, University of Illinois, Thomas A. Kochan, Massachusetts Institute of Technology, and John Thomas Delaney, Columbia University, editors

This important book analyzes the key issues of public-sector bargaining. It also takes an interesting look ahead to the problems that will confront policy makers and practitioners in the years to come.

1988 0-669-12893-7 640 pages $19.95 (paper)
0-669-17125-5 $45.00 (cloth)

NEW TEXT

Collective Bargaining in American Industry

Contemporary Perspectives and Future Directions

David B. Lipsky, Cornell University, and Clifford B. Donn, LeMoyne College, editors

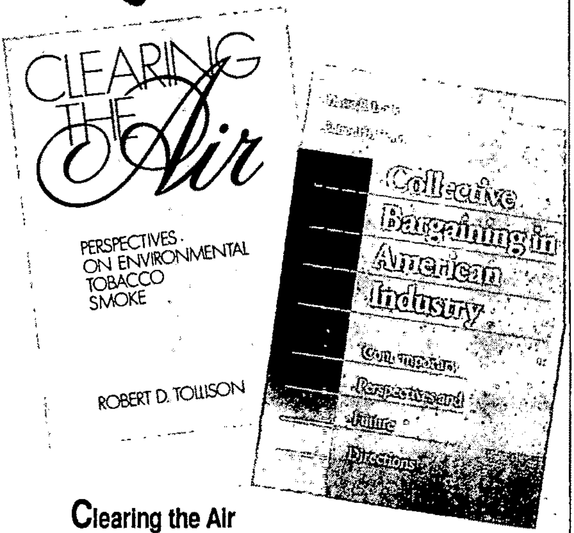
A dynamic look at a cross-section of American employers and employees caught in a swiftly-changing economic system.

1988 0-669-12595-4 304 pages $16.95 (paper)
0-669-12594-6 $37.00 (cloth)

DC Heath
A Raytheon Company

LEXINGTON BOOKS
125 Spring Street
Lexington, MA 02173

or call toll free: 1-800-235-3565


CLEARING THE Air
PERSPECTIVES ON ENVIRONMENTAL TOBACCO SMOKE
ROBERT D. TOLLISON
Collective Bargaining in American Industry
Contemporary Perspectives and Future Directions

Clearing the Air

Perspectives on Environmental Tobacco Smoke

Robert D. Tollison, George Mason University

Environmental tobacco smoke (ETS) stimulates many complex issues for debate and analysis, including economic incentives and the appropriate public policy toward ETS. This unique volume presents the ETS issue from several perspectives ranging from health science to economic.

1988 0-669-18007-6 160 pages $27.50 (cloth)

Smoking and the State

Social Costs, Rent Seeking, and Public Policy

Robert D. Tollison and Richard E. Wagner, George Mason University

Smoking: who pays for its costs? After careful analysis, the authors conclude that there is no basis for the claim that the economic costs of smoking are imposed upon nonsmokers, or that smokers impose uncompensated costs on nonsmokers from the possible consequences of environomental tobacco smoke. Their findings are a significant contribution to the ongoing smoking controversy.

1988 0-669-17100-X 144 pages $25.00 (cloth)

Houghton Mifflin 1988

Cases in Financial Management
Third Edition

Jerry A. Viscione and **George A. Aragon**
Both of Boston College

About 576 pages • hardcover • Instructor's Manual • Just published

The revised Viscione/Aragon text presents a wide range of cases that exemplifies the theories, policies, and techniques of modern business finance. The authors examine all sizes and types of firms, including nonprofit organizations. The cases are written at three different levels—from single solution to open-ended—making the text perfect for the first course in case analysis. Helpful notes on how to prepare a case are also provided. The Instructor's Manual provides complete solutions to all case questions.

Essentials of Business Statistics
Second Edition

Wayne W. Daniel
Georgia State University

426 pages • hardcover • Study Guide • Instructor's Resource Manual with Tests
Solutions Manual • MicroTest: Computerized Test Bank • Call-in Test Service
GPA: Grade Performance Analyzer
STAT+: Statistical Software Package
DATA+: Computerized Data Base to accompany STAT+ • Just published

Daniel's comprehensive text is designed specifically to meet the demands of the one-term business statistics course. The new edition has been completely revised based on recommendations from statistics instructors nationwide. As a result, the text presents all of the commonly taught statistical topics. Worked-out examples and review questions immediately reinforce presentation of new concepts and also appear at ends of chapters.

The text and accompanying ancillary package encourage and facilitate computer use by incorporating many data sets, and by providing exercises specifically designed for computer use. Now accompanying this edition, and free to adopters, is STAT+ statistical software for the IBM® PC.

Business Statistics: Basic Concepts and Methodology
Fourth Edition

Wayne W. Daniel and **James C. Terrell**
Both of Georgia State University

753 pages • hardcover • 1986
Complete ancillary package

Comparative Economic Systems
Second Edition

Paul R. Gregory
University of Houston, College Park
Robert C. Stuart, Rutgers—
The State University of New Jersey

575 pages • hardcover • Instructor's Manual with Test Items • 1985

Gregory and Stuart's Second Edition offers full, balanced treatment of the theories of capitalism and socialism, illuminated with major case studies of systems in the United States, the Soviet Union, and Yugoslavia.

Domestic Transportation: Practice, Theory, and Policy
Fifth Edition

Roy J. Sampson, University of Oregon
Martin T. Farris and **David L. Shrock**
Both of Arizona State University

640 pages • hardcover • Instructor's Manual • 1985

A best seller in the field, *Domestic Transportation* integrates application with theory and policy, and is therefore suitable for both business administration and economics students.

For adoption consideration, request examination copies from your regional Houghton Mifflin office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
101 Campus Dr., Princeton, NJ 08540

Announcing a New Journal:

INTERNATIONAL ECONOMIC JOURNAL

VOLUME 1 WINTER 1987 NUMBER 4

Unemployment in Developing Countries: Economist's Models in Light of Evidence from Sri Lanka	Paul Glewwe
Hamiltonian Approaches to the Analysis of External Debt	Changhee Chae
Effects of Rising Price of Gold on the LDC's Demand for International Reserves	Mohsen Bahmani-Oskooee
Power Structure and Rent Competition	Yoo Soo Hong
Tax Policy and Corporate Financial Structure in Korea	Kwang Choi
The Intertemporal Substitution Hypothesis: A Quantitative Reassessment	Jisoon Lee
Optimal Policy Intervention to Reduce Import Dependence	R. S. Nettle, M. Britten-Jones and K. Anderson

International Economic Journal is published quarterly in English by the Korea International Economic Association. The personal subscription price of a volume (which includes postage) is $30.00 per year. The institutional subscription price is $40.00 per year. Cheques should be made payable to the International Economic Journal and sent directly to the Managing Editor.

MANUSCRIPTS SUBMITTED *in Asia for publication should be sent to the Managing Editor, Professor Wontack Hong, Department of International Economics, College of Social Sciences, Seoul University, Seoul 151, Korea. Manuscripts from countries outside Asia should be sent to the Co-Editor, Professor Young Chin Kim, Department of Economics, Northern Illinois University, DeKalb, Illinois 60115, U.S.A.*

ECONOMETRIC SOFTWARE FROM TSP INTL

Now available for PCs and mainframe computers

TSP Version 4.1: with Probit, Logit, Tobit, sample selection, general maximum likelihood estimation, and robust standard errors for all procedures. A complete programming language for econometricians and data analysts in use at over 1000 sites worldwide, it includes regression, nonlinear simultaneous equation estimation, time series models, and many other features.

The PC version is interactive and identical to the mainframe version. Databanks and the saving and restoring of workspaces are now supported. PC TSP requires 512K RAM, a math chip (8087 or 80287), hard disk recommended.

For mainframes only

RATS Version 2.0: identical to the popular PC version. We are the authorized distributor of the mainframe version.

For more info, write or call (415) 326-1927
TSP International • PO Box 61015 • Palo Alto, CA 94306

DRYDENOMICS–

Revised for 1988!

ECONOMICS, 5/E

Edwin G. Dolan, *George Mason University,* and David E. Lindsey, *Federal Reserve System*

MACROECONOMICS, 5/E

Edwin G. Dolan, *George Mason University,* and David E. Lindsey, *Federal Reserve System*

MICROECONOMICS, 5/E

Edwin G. Dolan, *George Mason University*

The new edition of this popular principles of economics text successfully combines the most current information on theory and policy issues with the traditional core topics of economics. Features: new one-model option for teaching macro theory using AS/AD as the core model • greater emphasis on cases within chapters • integrated coverage of international economics • new perspectives in microeconomics focus on public choice theory, entrepreneurship, and information.

The full text is also available in two separate macro and micro paperbacks, each with its own extensive ancillary package.

New!

MONEY, BANKING AND MONETARY POLICY

Colin D. Campbell, *Dartmouth College,* Rosemary G. Campbell, and Edwin G. Dolan, *George Mason University*

Up-to-date coverage of banking, the Federal Reserve System and its policy with emphasis on recent developments and trends in the field.

Revised!

THE PRICE SYSTEM AND RESOURCE ALLOCATION, 10/E

Ross D. Eckert, *Claremont McKenna College,* and
Richard H. Leftwich, *Leftwich Associates*

This classic text continues to link modern microeconomic theory to real-world applications based on actual world events that are of interest to students *and* within their experience.

New!

COMPARATIVE ECONOMIC SYSTEMS

H. Stephen Gardner, *Baylor University*

A survey of the operation and performance of the economic systems in a wide range of capitalist, socialist, and developing countries. Special attention is paid to the problems involved in making international statistical comparisons.

D34-7AER-12

Revised!

MONEY AND BANKING: A MARKET-ORIENTED APPROACH, 3/E

Ivan C. Johnson and William W. Roberts,
both of California State University—Northridge

Depicts the forces that continually operate in financial markets and shows how financial institutions influence resource allocation in the aggregate economy.

Revised!

AMERICAN ECONOMIC HISTORY, 2/E

Robert C. Puth, *University of New Hampshire*

Chronicles the impressive growth of the economy from the 17th century to the present and analyzes why it developed the way it did.

New!

THE WORLD ECONOMY: TRADE AND FINANCE

Beth V. Yarbrough and Robert M. Yarbrough,
both of Amherst College

Integrates modern theoretical developments with recent world events and current policy debates. Goes beyond the usual Keynesian model to analyze price and output effects within an aggregate real demand/aggregate real supply framework.

The Best of 1987:

PUBLIC FINANCE: A CONTEMPORARY APPLICATION OF THEORY TO POLICY, 2/E,

Tax-Update Edition
David N. Hyman, *North Carolina State University*

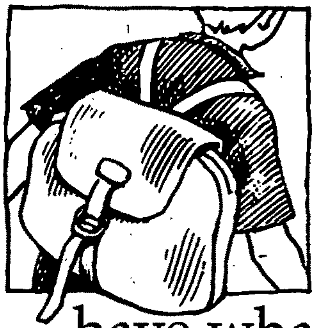
INTERMEDIATE MICROECONOMICS AND ITS APPLICATION, 4/E

Walter Nicholson, *Amherst College*

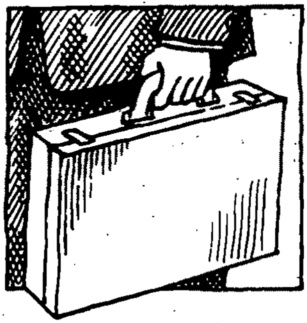
MANAGERIAL ECONOMICS, 5/E

James L Pappas, *University of South Florida,* and Mark Hirschey,
Rice University

THE DRYDEN PRESS
111 Fifth Avenue, New York, NY 10003



Do the books your students carry to class have what they'll need...



when they're carrying this?

Give your students an education they'll carry into their professional lives with these outstanding Harper & Row economics texts.

New for 1988

Miller
Economics Today, 6/e

Lipsey/Steiner/Purvis
Economics, 8/e
Now available in separate micro and macro volumes

Justice/Burton
Economics
An Introduction

Allen
Managerial Economics

Hamermesh/Rees
The Economics of Work and Pay, 4/e

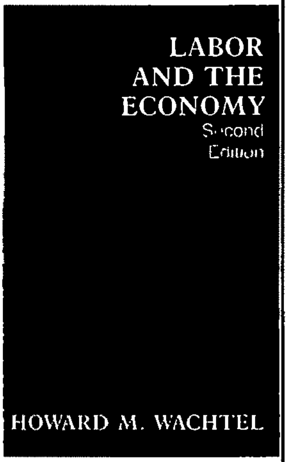
Markusen/Melvin
The Theory of International Trade

Stop by our booth to see how Harper & Row texts can give your students a lasting education, or call 1-800-223-4190.

1817 **Harper & Row**

HBJ
COMMUNIQUE

QUALITY AND SELECTION, AVAILABLE FROM HBJ



INTERMEDIATE MICROECONOMICS
ROBERT W. CLOWER
PHILIP E. GRAVES
ROBERT L. SEXTON
Hardcover, 544 pages, 1988

Links microeconomics and macroeconomics, emphasizing the verbal and intuitive presentation of topics.

THE EVOLUTION OF ECONOMIC THOUGHT
Fourth Edition
JACOB OSER
STANLEY L. BRUE
Hardcover, 544 pages, 1988

Revised and completely updated edition of a textbook noted for its clarity, scholarship, and balance.

LABOR AND THE ECONOMY
Second Edition
HOWARD M. WACHTEL
Hardcover, 544 pages, 1988

A thorough updating of this comprehensive textbook, integrating theory and institutions in the context of policy debates.

MACROECONOMICS
An Introduction to Advanced Methods
WILLIAM M. SCARTH
Hardcover, 288 pages, 1988

A concise, modern, new textbook that makes theoretical and mathematical debates comprehensible within an unbiased, thorough, research-oriented presentation.

HBJ

HARCOURT BRACE JOVANOVICH, Inc.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648

BLACKWELL

The Limits to Rational Expectations

HASHEM PESARAN

This is a critical assessment of the rational expectations hypothesis from theoretical as well as empirical perspectives. Covering a broad range of topics in the literature, it brings together in one place for the first time the diverse econometric contributions recently made in the areas of identification, estimation, and hypothesis testing in a wide variety of rational expectations models.
336 pages, **$39.95** (0 631 14036 0)

A Unified Theory of Estimation & Inference for Nonlinear Dynamic Models

A. RONALD GALLANT and HALBERT WHITE

Building on the great advances that have been made by statisticians and econometricians in the last forty years, this study presents the first unified theory of estimation and inference applicable to a wide variety of econometric estimators of the parameters of time-dependant heterogenous economic phenomena.
164 pages, **$34.95** (0 631 15765 4)

Production and Decision Theory under Uncertainty

KARL AIGINGER

This book contrasts theoretical models dealing with optimal decisions of firms under uncertainty with the empirical behaviour of firms in industrialized countries during the seventies and eighties. The analysis focuses particularly on the question of whether firms will produce and invest more, the same, or less under conditions of uncertainty.
Series: **Uncertainty & Expectations in Economics**
256 pages, **$49.95** (0 631 14792 6)

Modelling Trade and Trade Policy

JAN I. HAALAND and VICTOR D. NORMAN

An important collection of essays covering an issue of crucial topical interest to a wide range of economists: trade, and the international trading relations between countries and economic groupings. The papers were originally presented at the International Trade Conference held in Bergen in May 1986.
Series: **Scandanavian Journal of Economics**
180 pages, paperback **$22.50** (0 631 15943 6)

Basil Blackwell
108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016.
Toll-free ordering: 1-800-638-3030 (USA)

NEW BOOKS IN ECONOMICS
FROM
GREENWOOD PRESS, PRAEGER PUBLISHERS, AND QUORUM BOOKS

SOCIOECONOMIC ASPECTS OF RENEWABLE ENERGY TECHNOLOGIES
Edited by **Ramesh Bhatia** and **Armand Pereira** in collaboration with **Madhura Swaminathan**
This book addresses oversights in the current literature on energy programs in the developing countries by focusing on renewable energy technologies and topics. Among the subjects covered are ethanol production, renewable energy technologies for irrigation, and micro-economic evaluation of renewable energy programs.
March 1988. ISBN 0-275-92851-9. $46.00 est.
SUPPLY-SIDE PORTFOLIO STRATEGIES
Edited by **Victor A. Canto** and **Arthur B. Laffer**
Written by the internationally recognized originator of the "Laffer Curve" and other leading supply-side strategists, this book highlights the significance of incentive economics and its investment applications in today's volatile and uncertain economic climate.
February 1988. ISBN 0-89930-286-6. $35.00
THE MAKING OF UNITED STATES INTERNATIONAL ECONOMIC POLICY
Principles, Problems, and Proposals for Reform. Third Edition.
by **Stephen D. Cohen**
A revised and updated edition of the definitive study of how U.S. international economic policy is formulated and how the decision-making process functions. Major changes from previous editions include a fuller development of the definition of international economic policy and of the ideology underlying American policy in that arena; an analysis of the impact of private sector lobbying on public policy; and a description and analysis of the Reagan administration's international economic policies.
February 1988. ISBN 0-275-92334-7. $45.00 (hardbound) ISBN 0-275-92292-8. $17.95 (paperback)
REFEREED AND NONREFEREED ECONOMIC JOURNALS
A Guide for Publishing Opportunities
Compiled by **A. Carolyn Miller** and **Victoria A. Punsalan** with the assistance of **Kenneth G. Rohm**
This directory is designed to aid prospective authors seeking publication opportunities in economic journals. In addition to providing manuscript submission information for over two hundred academic and professional economic journals accepting manuscripts in English, it includes specific information on reviewing practices and clearly identifies refereed and nonrefereed journals.
January 1988. ISBN 0-313-25857-0. $39.95
PROPERTY, POWER, AND PUBLIC CHOICE
An Inquiry Into Law and Economics. Second Edition.
by **A. Allan Schmid**
A revised and updated edition of a work hailed by *The Journal of Economic Education* as "a benchmark book that provides the reader with a solid base of what constitutes the subject matter of law and economics."
1987. ISBN 0-275-92797-0. $49.95 (hardbound) ISBN 0-275-92828-4. $23.95 (paperback)
THREATS OF QUOTAS IN INTERNATIONAL TRADE
Their Effect on the Exporting Country
by **Gerard Lawrence Stockhausen**
This volume provides five case studies of threats and protection in particular industries as a basis for discussion of the political economy of quota threats.
March 1988. ISBN 0-313-25785-X. $37.95
STALEMATE
Political Economic Origins of Supply-Side Policy
by **Howard A. Winant**
Winant examines the politics of economic policy formation in the postwar United States. He critically reinterprets political sociological approaches to the contemporary state/economy relationship, compares mainstream and radical accounts of the policy making process, and proposes a synthesis between pluralist, elite, and class-based approaches.
January 1988. ISBN 0-275-92706-3. $39.95

Greenwood Press/Praeger Publishers/Quorum Books
Divisions and Imprints of Greenwood Press, Inc.
88 Post Road West, P.O. Box 5007, Westport, Conn. 06881 (203) 226-3571

New from Cambridge University Press

Politics, Work, and Daily Life in the USSR

A Survey of Former Citizens

James R. Millar, Editor

"Professor Millar and his colleagues have rendered the field of Soviet studies a signal service by systematically investigating the attitudes, values, and life histories of a large sample of 'Third Wave' emigres. Their findings help to clarify some of the most hotly disputed analytical issues in the field. All of the contributions to this volume will repay careful reading and some of them will undoubtedly become classics in the field."

Jeremy R. Azrael, The RAND Corporation

Cloth $49.50   Paper $16.95

Job Rights in the Soviet Union: Their Consequences

David Granick

In this pioneering study of the labor market in the contemporary Soviet Union, Professor Granick shows how its distinctive features of low unemployment and the very low dismissal rates are not the product of high-boom economic overheating in the capitalist sense but the deliberate consequence of specific planning objectives.

$49.50

The Americanisation of West German Industry, 1945-1973

Volker R. Berghahn

The author looks at "Americanizing" influences in West Germany in the fields of market behavior, industrial relations, lobbyism, economic planning and European integration and suggests entirely new explanations for the postwar success of German industry.

$34.50

World Coal

Economics, Policies and Prospects

Richard L. Gordon

Richard Gordon provides a comprehensive study of coal in the energy market. He examines the economics of coal production, the reasons for its past decline and its prospects for the future, and shows that while coal will probably never recover the market it has already lost, it will continue to remain important because of its competitiveness in the generation of electricity.

Cambridge Energy Studies

$39.50

The Great Depression

Delayed Recovery and Economic Change in America, 1929-1939

Michael A. Bernstein

By examining the uneven fate of manufacturing industries during the thirties, Michael Bernstein presents a powerful new interpretation of the Great Depression. He provides a more detailed picture of what happened to the American economy in the thirties that was so different from previous downturns.

$29.95

The Long-Run Behavior of the Velocity of Circulation

The International Evidence

Michael D. Bordo and Lars Jonung

This study develops and provides empirical evidence for a coherent framework to explain the secular behavior of the income velocity of money. Existing theories can explain each section of the velocity curve but the overall pattern is not consistent with any one prevailing theory. In this book, Bordo and Jonung offer a comprehensive explanation for velocity's secular behavior.

$34.50

Advances in Economic Theory, Fifth World Congress

Truman F. Bewley, Editor

This volume with its focus on economic theory, and its companion volume on econometrics, contains papers delivered at the Fifth World Congress, held in 1985. It is designed to make material accessible to a general audience of economists.

Econometric Society Monographs 12

$47.50

Advances in Econometrics, Fifth World Congress

Volumes 1 and 2

Truman F. Bewley, Editor

These two volumes, with their focus on econometrics, contain papers delivered at the Fifth World Congress held in 1985. They include a variety of topics, comprising empirical and policy oriented subjects as well as theoretical and methodological ones.

Econometric Society Monographs 13 & 14

Volume 1 $34.50
Volume 2 $34.50

Contributions to Econometrics

John Denis Sargan

Esfandiar Maasoumi, Editor

These volumes bring together some of Denis Sargan's major contributions to econometrics and economics, including both previously published and unpublished papers.

Volume 1 about $44.50
Volume 2 about $44.50

At bookstores or order from

Cambridge University Press
32 East 57th Street, New York, NY 10022.
Cambridge toll-free numbers for orders only:
800-872-7423, outside NY State. 800-227-0247, NY State only.
MasterCard and Visa accepted.

Developments in Economics

Journal of the Japanese and International Economies

An International Journal

Editorial Board

Masahiko Aoki
University of Kyoto and Stanford University

Michihiro Ohyama
Keio University

Koichi Hamada
Yale University

Masahiro Okuno-Fujiwara
University of Tokyo

Devoted to Academic Analyses of the Japanese Economy and Its Interdependence with Other National Economies
Published in Cooperation with the Tokyo Center for Economic Research

Volume 2 (1988), 4 issues ISSN 0889-1583
In the U.S.A. and Canada: $100.00
All other countries: $112.00

Journal of Economic Theory

Editor

Karl Shell
Cornell University, Ithaca, New York

Economic Theory Emphasizing the Theoretical Analysis of Economic Models Including the Study of Related Mathematical Techniques

Volumes 44–46 (1988), 6 issues
ISSN 0022-0531
In the U.S.A. and Canada: $330.00
All other countries: $394.00

Sample copies and privileged personal subscription rate information are available. For more information, please write or call:

ACADEMIC PRESS, INC.
Journal Promotion Department
1250 Sixth Avenue
San Diego, CA 92101 U.S.A.
(619) 230-1840

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
San Diego New York Boston London Sydney Tokyo Toronto

S8007B

You are cordially invited to become a

Charter Subscriber

to

CHANCE

NEW DIRECTIONS FOR STATISTICS AND COMPUTING

Like no other statistics publication, CHANCE: NEW DIRECTIONS FOR STATISTICS AND COMPUTING is not only informative and immediately useful... it's entertaining!

CHANCE captures the excitement, innovation, and fresh perspective that statistics can yield. While enjoying **CHANCE**, you learn of the newest ideas and applications in statistics—and how these are intertwined with advances in computer technology. Technical, practical information is presented with verve and good writing, well illustrated for greater comprehensibility and attractiveness. Each issue is enlivened with quotes from the press, musings on statistics and its environment, cartoons, puzzles, contests, and poetry, satire, and other styles of presentation, not usually found in statistical publications.

Stimulate your thinking on innovative uses of statistics in the widest range of applications—

Our authors are statisticians and others applying statistical ideas and methods to real-world problems. Read **CHANCE** for reports on advances resulting from the use of statistical approaches in the social, biological, physical, and medical sciences. Learn more about the growing importance of statistical applications in industry—quality control and measurements of productivity, for instance—and in business, especially marketing analysis. But we don't overlook the traditional strongholds of statistics: you'll see plenty about new approaches to conducting national censuses and measuring economic performance. You may be amused, often excited, by what you read on the uses and misuses of statistics in major public-policy debates, by discussions of the future of the statistics profession, as well as by notes from the history of the field, its ideas, and its practitioners.

Unique focus on computers and computer technology as they are intertwined with the practice of statistics—

In **CHANCE**, in every issue, you will find reviews, discussions, and comparisons of the software available to you now. And you'll read progress reports as new programs are developed to address particular applications. You'll receive our advice on good buys and what to avoid. Also featured are regular columns devoted to improving your efficiency in personal computing and to communicating more effectively through computer-generated graphics. **CHANCE** gathers together the information on computer technology—hardware and software—as it affects your work.

As a Charter Subscriber, you are entitled to examine the first issue with no obligation.
You risk absolutely nothing by taking a look!

RSVP

A Special Invitation to Become a Charter Subscriber

RSVP

To accept this special invitation, return this coupon today or call toll-free 1 (800) 526-7254.

IMPORTANT GUARANTEE!

If you are not satisfied with **CHANCE: NEW DIRECTIONS FOR STATISTICS AND COMPUTING** at any time, you may cancel your subscription and receive a full refund on all unmailed issues, no explanation necessary.

☐ **YES**, I'd like to become a charter subscriber to **CHANCE: NEW DIRECTIONS FOR STATISTICS AND COMPUTING** and try the first issue with no obligation!

Please rush it to me and enter my charter subscription. I will review this issue and, if I wish, I will pay your subscription invoice of $25.95 to receive three additional issues. If I am not fully satisfied, I will return the invoice marked "no thank you." In that case, the first issue is mine to keep and I will owe nothing.

Send my issue to: ____________________

____

____

____________________ **107F**

Special personal rate only. Institutional rate: $49.95 (4 issues). Offer valid only in North America.

SPRINGER-VERLAG NEW YORK, INC. P.O. Box 2485, Secaucus, NJ 07094

New from the Consortium of Social Science Associations....

The first comprehensive

Guide to Federal Funding for Social Scientists

Prepared by the Consortium of Social Science Associations (COSSA), a Washington advocacy group serving the major professional societies in the social and behavioral sciences. Susan D. Quarles, editor.

The federal government is a major supporter of research in the social and behavioral sciences, but until now, no single, multidisciplinary directory has been available to guide researchers through the complexities of government funding in these fields.

COSSA's inclusive new *Guide to Federal Funding* describes over 300 federal programs in impressive detail, including funding priorities, application guidelines, and examples of funded research. Introductory essays describe the organization of social science funding and offer inside views of federal funding practices and contract research.

For anyone who needs to know the ins-and-outs of government funding in the social sciences and related fields, COSSA's *Guide* will be an essential new resource.

Published by the Russell Sage Foundation

512 pages ISBN 0-87154-699-X Paperback

Mail orders to:

Consortium of Social Science
Associations
1625 I St., NW, Suite 911
Washington, D.C. 20006

Orders will be filled when books are available, in June. Please allow 3-4 weeks for delivery.

Please send ______ copies of COSSA's *Guide to Federal Funding for Social Scientists* (81-0699X) at the following price:

☐ $24.95 (for libraries and institutions)

☐ **$14.95 (special price for members of AEA)**

☐ $19.95 (for other individuals)

Enclosed is a check, money order, or purchase order in the amount of $ ________. Publisher pays postage on prepaid orders; New York residents, please add sales tax.

____

Name (please print)

____

Address

____

CSWEP

The Committee on the Status of Women in the Economics Profession

CSWEP, an arm of the American Economic Association (AEA), publishes a thrice-yearly newsletter to let you know all the news—good or bad—about women in the economics profession.

The newsletter carries items about:

job opportunities

research funds, scholarships, and fellowships

conferences and calls for papers

relevant developments in government, industry, and academia

women's progress within the profession

new research by and about women

CSWEP represents women's point of view in the committee work of the AEA and makes an annual report to the AEA on the status of women economists.

CSWEP is a presence at the annual meetings of the AEA and the regional economics associations. It sponsors technical sessions on topics of particular interest to women and hosts social events.

CSWEP publishes a ROSTER OF WOMEN ECONOMISTS for purposes of communication and job placement. Dues-paying members receive the latest roster, which lists women economists by where they teach or work, by their specialty in economics, and by city, as well as alphabetically.

CSWEP provides information and encouragement to students thinking about graduate work or careers in economics.

To become a dues-paying member of **CSWEP**
send this with a check for $20 (tax deductible) made out to **CSWEP** to:
CSWEP, c/o Joan G. Haworth
4901 Tower Court, Tallahassee, FL 32303

NAME____

MAILING ADDRESS________________________________

CITY, STATE, ZIP__________________________________

Check here if currently an AEA member ☐

Full-time students may join free:____________________________
(Name of Student's Institution)

Computer Access to Articles in the JEL Subject Index

Online computer access to the *JEL* and *Index of Economic Articles* database of journal articles is currently available through DIALOG Information Retrieval Service. DIALOG file 139 *(Economic Literature Index)* contains complete bibliographic citations to articles from the nearly 300 journals listed in the quarterly *JEL* issues from 1969 through the current issue. The abstracts published in *JEL* since June 1984 are also available as part of the full bibliographic record. The *Economic Literature Index* also includes citations to articles in the 1979 and 1980 collective volumes (collected papers, proceedings, etc.) for the *Index* database; other years will be added as soon as completed. The file may be searched using free-text searching techniques or author, journal, title, geographic area, date, and other descriptors, including descriptor codes based on the *Index's* four-digit subject classification numbers. (For a complete description of the *Economic Literature Index* with search examples and suggestions for searching techniques, see the article "Online Information Retrieval for Economists—The Economic Literature Index," in the December 1985 issue of the *Journal of Economic Literature.)*

Access Options:

- **DIALOG** offers a variety of contract choices, including the option (for a low annual fee) to pay for only what you use. Most university libraries already subscribe to DIALOG. For information on the DIALOG service, contact your librarian or write to or call: DIALOG Information Services, Inc., Marketing Department, 3460 Hillview Avenue, Palo Alto, California 94304 (800-3-DIALOG or 800-334-2564).

- **Knowledge Index**, a DIALOG service available after 6 p.m. and on weekends, may be accessed at the low rate of $24/hour, charged to a major credit card. A one time start-up fee of $35.00 buys 2 hours free time during the first month after log-on. Call 800-3-DIALOG for information.

- **EasyNet**, a gateway service, provides menus to guide the untrained user through database searches in DIALOG and other databases. For information, call 1-800-841-9553 or dial up **EasyNet** on your terminal (1-800-EASYNET) and pay for your search by credit card.

Classroom Instruction:

- DIALOG's Classroom Instruction Program, available at a special rate of $15/connect hour to academic institutions for supervised instruction, permits teachers to incorporate online bibliographic searching in their courses. For information, contact DIALOG or your librarian.

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class):	$15.00, regular AEA members and institutions $ 7.50, junior members of AEA
All other countries (air mail):	$22.50, regular AEA members and institutions $15.00, junior members of AEA

Please begin my issues with:
☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name ____
First Middle Last

Address ____

____
City State/Country Zip/Postal Code

Check one:

☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

Journal of International Economic Integration

Solicits Papers to Compete for the Annual Daeyang Prize in Economics ($7,000) and Welcomes Subscriptions by Interested Parties

Current issues include

Bela Balassa, *Japanese Trade Policies Towards Developing Countries.*

Basant K. Kapur, *Open-Economy Response to a Terms of Trade Shock in a Growth Context.*

Norman C. Miller, *A Gerneral Approach to the Balance of Payments and Exchange Rates.*

Leonard F.S. Wang, *Product Market Imperfections and Customs Unions theory.*

Gene M. Grossman, *The Employment and Wage Effects of Import Competition in the United States.*

Wilfred J. Ethier and Ronald D. Fischer, *The New Protectionism*

Joshua Aizenman, *Inflation, Tariffs and Tax Enforcement Costs*

Chung H. Lee and Seiji Naya, *The Internationalization of U.S. Service Industries and Its Implications for Developing Countries.*

The Journal of International Economic Integration is published biannually (Spring and Autumn) by the Institute for International Economics, King Sejong University, Seoul, Korea.
The purpose of the Journal is to support and encourage research in the area of international trade, international finance and other related economic issues that include general professional interest in international economic affairs. Welcoming both theoretical and empirical analyses in international economics, the Journal is strongly interested in the issues of the international economic cooperation.

- The Journal welcomes unsolicited manuscripts, which will be considered for publication by the Editorial Board.
- From papers selected for publication, the Prize committee will choose the best manuscript(s) to receive the $7,000 Daeyang Prize. The winner of the prize is announced in the Spring issue every year.
- The manuscripts should be accompanied by an abstract of no more than 100 words and a brief curriculum vitae containing the author's academic career. All submissions should be typewritten, double-spaced, in English with footnotes, references, figures, tables and any other illustrative material on separate sheets.
- Three copies of the manuscript and all accompanying material should be submitted to the following address by October 31, 1988 for consideration for 1989 publication.
- For subscriptions to the Journal ($20 per year for individuals, $30 per year for institutions), send a check or money order payable to King Sejong University to the following address.

Institute for International Economics
King Sejong University
Seongdong-Ku, Seoul, Korea